Sunk costs, extensive R&D subsidies and permanent inducement effects

Working Paper Series

#2012-029

Sunk costs, extensive R&D subsidies and permanent inducement effects

Pere Arqué-Castells and Pierre Mohnen

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

Sunk costs, extensive R&D subsidies and permanent inducement effects

Pere Arqué-Castells[a]
Universitat de Barcelona & Institut d'Economia de Barcelona, Dpt. d'Economia Pública, Economia Política i Economia Espanyola, Av. Diagonal 690, Torre 4, Planta 2, 08034 Barcelona, Spain

Pierre Mohnen[b]
University of Maastricht, UNU-MERIT, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

Abstract
We study whether there is scope for using subsidies to smooth out barriers to R&D performance and expand the share of R&D firms in Spain. We consider a dynamic model with sunk entry costs in which firms' optimal participation strategy is defined in terms of two subsidy thresholds that characterize entry and continuation. We compute the subsidy thresholds from the estimates of a dynamic panel data type-2 tobit model for an unbalanced panel of about 2,000 Spanish manufacturing firms. The results suggest that "extensive" subsidies are a feasible and efficient tool for expanding the share of R&D firms.

Keywords: R&D, Persistence, Subsidies, Dynamic models
JEL Codes: H2, O2, C1, D2

---

[a] Tel.: +34 934034729; fax: +34 934037242. E-mail address: pere_arque@ub.edu.
[b] Tel.: +31 43 388 4464; fax: +31 43 388 4905; E-mail address: mohnen@merit.unu.edu

We are grateful to José García-Quevedo, Emmanuel Duguet, Elena Huergo, Georg Licht, Jarle Møen, Juan Sanchís-LLopís and Alejandro Esteller-Moré for their helpful comments. We also thank participants at the Workshop on R&D Policy Evaluation (Paris, 2011), XXXVI Simposio de Economía Española 2010, EARIE 2010 conference, ESSID 2009, Zvi Griliches Summer School 2009, DRUID 2009 conference and internal IEB, UNU-MERIT, Universidad de Valencia and Université de Bordeaux seminars. Financial support from the Instituto de Estudios Fiscales, the Generalitat de Catalunya (SGR-2009-102 project) and the Spanish Ministry of Science and Innovation (ECO2010-16934) is gratefully acknowledged.

1. Introduction

All countries have R&D and innovation support programmes to spur growth by overcoming market failures. Such programmes comprise a wide range of tools including tax cuts, subsidies for performing R&D activities, the creation of technological laboratories or innovative clusters. Of all these forms of public support, subsidies are in most countries the principal tool of public intervention.

Subsidy policies aimed at enhancing the overall R&D expenditure of a given country can follow two different courses of action. On the one hand, they can act on the intensive margin, seeking to promote the R&D effort of regular R&D performers. On the other hand, they can act on the extensive margin, seeking to expand the base of R&D performers. Traditionally, subsidy policies have followed the first course of action (see Blanes and Busom, 2004; Aschhoff, 2008; Huergo and Trenado, 2010). Similarly, and possibly as a consequence, most of the research on R&D subsidies has focused on the intensive margin too (see the surveys by Klette et al., 2000; David et al., 2000; and Garcia-Quevedo, 2004).

This lack of interest in "extensive" subsidies is hard to understand for one main reason. It is only those countries with a substantial share of R&D firms that achieve high R&D intensities (see Figure 1). So even if the final goal is to increase R&D intensity it must necessarily be achieved through expanding the number of R&D firms. Countries acting on the intensive margin alone are unlikely to meet the European Commission's target of spending 3% of GDP on R&D.

[INSERT FIGURE 1]

Our goal is to study whether there is scope for using "extensive" subsidies to expand the share of R&D firms of a given country. At the limit all firms could be subsidized but this would be costly and not necessarily welfare enhancing. So a first step is to define which circumstances if any justify the use of "extensive" subsidies. Our justification is related with the existence of sunk entry costs in R&D activities. Becoming an R&D-performing firm is costly as it often requires setting up a new department, hiring and training researchers and investing in machinery. These outlays are generally nonrecoverable and can be considered as sunk costs. As a result of the existence of sunk entry costs, some firms are likely to need subsidies to start but not to continue performing R&D. We defend that "extensive" subsidies should essentially be used to smooth out the sunk entry costs of these firms.

We set out to detect whether such a group of firms exists in Spain, a low R&D-intensity country with a small share of R&D firms. We also aim to quantify the costs (total amount of subsidies) and benefits (total R&D stock generated) derived from inducing this group of firms into R&D. In short, we ask ourselves how far Spain can progress along the linear fit of Figure 1 with a policy based on "extensive" subsidies.

To this end, we consider a dynamic model with sunk entry costs in which firms decide whether to start, continue or stop performing R&D on the grounds of the subsidy coverage (share of to-be-made R&D expenditures) they expect to receive (our model can be seen as a dynamic version of González et al. (2005)). Firms' optimal participation strategy is defined in terms of two subsidy (or R&D) thresholds that characterize entry and continuation. The entry threshold is larger than the continuation threshold owing to the fact that firms are in greater need of aid when they lack experience in R&D and sunk costs still need to be paid. Whenever the expected subsidy coverage is above the entry (resp. continuation) threshold, firms find it optimal to enter (resp. continue doing) R&D. Temporary subsidies above the entry threshold can lead to permanent R&D activity as long as the level of subsidies remains above the continuation threshold. Firms with positive entry thresholds and zero or negative continuation thresholds can be permanently induced into R&D by means of one-shot trigger subsidies.

Firms' optimal participation policy can be cast in terms of a type-2 tobit specification with dynamics in the selection equation where firms find it optimal to enter (resp. continue doing) R&D if optimal R&D expenditure is above the R&D entry (resp. continuation) threshold. To deal with selectivity in R&D performance we implement Raymond et al. (2010) random effects estimator, which follows Wooldridge (2005) in treating the unobserved individual effects and the endogeneity of the initial conditions. Given that our structural model satisfies the identification restrictions highlighted in

Nelson (1977) we are able to recover the R&D and subsidy thresholds for every single firm.

We estimate the model using an unbalanced panel of more than 2,000 Spanish manufacturing firms observed during the period 1998-2009. The dataset includes numerous entries into, and exits from, R&D and reports information on R&D spending. Somewhat unusually the dataset also contains information on both successful and rejected subsidy applicants, the latter information being crucial for identifying subsidies' inducement effects.

Subsidies are presumably endogenous as they are granted by agencies according to the effort and performance of firms. To deal with this problem we assume that firms react to subsidies expected in advance along the lines of González et al. (2005). However, we construct a slightly different measure of expected subsidies drawing on the information we have on subsidy applicants. This will enable us to control for fixed effects via the inclusion of Mundlak means which will ultimately result in a better identification of the subsidy parameters.

The paper leads to a series of interesting findings. First of all, expected subsidies significantly affect both R&D expenditure and the decision to perform R&D. In addition, there is true state dependence in the sense that firms that perform R&D in a given period are 37% more likely than those that do not to perform R&D in the next period. This result implies that there are two subsidy thresholds rather than one, which allows for permanent inducement effects. The subsidy thresholds are used to classify firms according to their dependence on subsidies for making the performance of R&D activities profitable. Interestingly, 10% of Spanish manufacturing firms are found to need subsidies when they lack any previous experience, but they can persist in R&D without them. We estimate that inducing this group of firms would cost €110 million, while the yearly R&D investments that would be triggered is estimated at €453 million and the R&D stock generated at €2,500 million in 15 years. "Extensive" subsidies would move Spain from its current position in Figure 1 to somewhere between Italy and Ireland.

Our paper is most closely related to González et al. (2005) who study the effectiveness of subsidies at stimulating R&D performance in a setting with fixed (but not sunk) costs. They find that subsidies can encourage non-R&D performing firms to start investing in R&D but, unlike us, they are unable to tell whether firms need different subsidy shares to start or to continue performing R&D. In our dynamic framework with sunk entry costs a firm's optimal participation strategy can be defined in terms of two (rather than one) subsidy thresholds characterising entry and continuation. This enhancement proves crucial as it allows us to detect the permanent inducement effects (which go unnoticed in a static setting) that make "extensive" subsidies a feasible and efficient tool to induce entry.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 develops the analytical framework. Section 4 presents the econometric modeling and discusses the main identifying assumptions. Section 5 outlines the empirical specification and presents the estimates as well as some robustness checks. Section 6 discusses the policy implications of our results regarding subsidy coverage thresholds for R&D entry and continuation, the extent of permanent inducement effects and the evaluation of the costs and benefits of an extensive R&D subsidy policy. Section 7 concludes.

2. Data

The dataset we use is the "Encuesta Sobre Estrategias Empresariales" (from now on ESEE)[1]. This survey gathers information from manufacturing firms operating in Spain employing more than nine workers. It is conducted on a yearly basis across twenty different sectors. The initial sampling undertaken in conducting the survey differentiated firms according to their size. While all firms employing more than 200 employees were required to participate, firms with between 10 and 200 employees were selected by stratified sampling (stratification across the twenty sectors of activity and four size intervals). Subsequently, all newly created firms with more than 200

[1] The ESEE (Survey on Firm Strategies) has been conducted since 1990 by the *Fundación SEPI* under the sponsorship of the Spanish Ministry of Industry.

employees together with a randomly selected sample of new firms with between 10 and 200 employees have been gradually incorporated.

The survey keeps track of the firms' technological activity and reports information on several measures of R&D performance including intramural expenditure, R&D contracted with external laboratories or research entities and technological imports. For our purposes, a firm is classified as an R&D performer whenever it reports having incurred expenditure in any of these categories excluding technological imports.[2]

In addition, the survey provides information on the R&D subsidies received by successful subsidy applicants. The subsidy variable we use considers the total quantity of aid granted by the various public agencies (primarily the national agency, CDTI, but also regional and European agencies). We can also identify rejected subsidy applicants from a question available in the ESEE since 1998 that asks firms whether they sought external financing without success[3]. Since the public sector is by far the main available source of external financing in Spain we can safely view firms claiming to have sought external R&D funding without success as rejected subsidy applicants[4]. This was confirmed by the technical director of the ESEE[5].

In this study, we use survey data from 1998 to 2009 [6]. The cleaned panel data sample comprises 14,283 observations corresponding to 2,621 firms observed over a varying number of years (see Table 1), 4,524 R&D observations, 1,585 R&D funding applications and 1,082 successful applications[7]. Approximately 2/3 of applications were accepted. This acceptance rate is in line with the figures found in other papers (see Takalo et al., forthcoming; Huergo and Trenado, 2010). Remarkably only 6% of the

[2] Our definition of R&D is consistent with the definition given in the Frascati Manual (OECD, 2002) definition.
[3] The exact question is "Did you search external R&D funding without success?".
[4] According to the PITEC, on average, 81% of Spanish firms' R&D expenditures are funded with own internal funds while 16.7% are funded with public funds (both from Spanish and European administrations) and only 2.3% come from other sources. So almost all external funding comes from the public sector.
[5] The technical director of the ESEE told us that their internal checks clearly suggest that the outcomes of the question "Did you search external R&D funding without success?" can be used to infer whether firms applied for subsidies without success.
[6] We do not use previous years because information on subsidy applicants, which is key to identifying the subsidies inducement effects, is only available since 1998.
[7] To obtain the cleaned dataset we have simply deleted data points for which relevant variables are missing. We have also deleted some observations with subsidies higher than R&D expenditures that are not consistent with our empirical modeling.

subsidies are granted to firms that did not perform R&D in the previous period. This suggests that subsidies are mainly targeted at active R&D firms and very rarely used to encourage entry into R&D.

[INSERT TABLE 1]

Table 2 shows the importance of having data on funding applications to study the subsidies inducement effects. While all successful applicants perform R&D, only 72% (63% of firms that continue plus 9% of entrants) of rejected applicants do so. Interestingly, 24% of rejected applicants fail to enter into R&D and 4% are forced to abandon R&D presumably due to the lack of financing. This group of rejected applicants would have performed R&D had it received subsidies. This suggests that subsidies do have some inducement effects.

[INSERT TABLE 2]

Table 3 provides an initial insight of the extent to which firms engage in R&D activities as well as of the stylized facts governing the assignment of subsidies to R&D performers. A marked stylized fact is that the proportion of R&D performers increases greatly with size. Whereas, in most years, only around 20% of firms with fewer than 200 workers perform R&D, this percentage rises to 70% when we consider firms with more than 200 workers. Similarly, the proportion of subsidized firms among R&D performers increases with firm size. Whereas only 10% to 25% of R&D performers with fewer than 200 workers enjoy subsidies, 25% to 39% of R&D firms with more than 200 workers receive subsidies. As for the subsidy coverage (ratio of subsidy to R&D expenditure), this adopts a mean value of 34% for firms with fewer than 200 workers, falling to 25% in the case of firms with more than 200 workers[8]. Hence, the proportion of subsidized R&D expenditure declines with firm size.

Interestingly, there is a sharp increase in the percentage of subsidized R&D performers from 2004 onwards coinciding with a change in the Spanish government (from conservative to socialist party). However, the average number of R&D firms and the

[8]These numbers are valid for the subsample of R&D firms only. The average subsidy coverage is much lower for the entire sample (see Table 4A in the Appendix).

average subsidy coverage remain unaltered. This confirms that the government sought to increase R&D expenditures by focusing on the intensive margin (i.e., subsidizing active R&D firms).

[INSERT TABLE 3]

Table 4 differentiates between stable and occasional R&D performers and provides more detail on the probability of a firm undertaking R&D and being granted a subsidy. It appears that stable R&D performance, understood as performing R&D during the whole sample period, is mainly observed in large firms and that it is quite uncommon among small firms. By contrast, occasional performance is more evenly distributed among firms of different sizes, being most common among medium-sized firms. If we focus solely on R&D performers, the probability of being granted a subsidy increases markedly with firm size and stable performance.

[INSERT TABLE 4]

3. Analytical framework

In this section we present a stylized analytical setting that illustrates how public subsidies modify firms' optimal R&D decisions (whether to perform R&D and how much to invest). We will then draw on this set up to build our empirical specification.

3.1. Demand

We consider a product-differentiated market with monopolistic competition in which firms produce a single type of each variety of good. These varieties are symmetrically differentiated, with common elasticity of substitution $\sigma > 1$ between any two of them. The demand for firm i's output, q_{it}, is generated by a representative consumer that spends a fixed amount of income Y on the products of the industry. The utility function is of the Dixit-Stiglitz (1977) type augmented to accommodate the consumer's valuation[9]:

[9] The consumer's valuation is introduced in line with Melitz (2000).

$$U(\Lambda_{it} q_{it}) = \left(\sum_i (\Lambda_{it} q_{it})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where $U(\cdot)$ is assumed to be differentiable and quasi-concave and Λ_{it} represents the consumer's valuation of firm i product. Utility maximization gives demands of the form:

$$q_{it}(p_{it}, \Lambda_{it}) = z_t\, p_{it}^{-\sigma} \Lambda_{it}^{\sigma-1} \tag{2}$$

where $p_{it} = \overline{p}_{it} / \widetilde{p}_{it}$ is the deflated price, $\overline{p}_{it}$ is the nominal price, $\widetilde{p}_{it} = \left[\sum \left(\overline{p}_{it} / \Lambda_{it} \right)^{(1-\sigma)} \right]^{1/1-\sigma}$ is a quality-adjusted price index (a price deflator) and $z_{it} = Y_t / \widetilde{p}_{it}$ captures exogenous demand shifters. Firms are considered too small relative to the industry to influence the aggregate $\widetilde{p}$-index[10] and so competitive interaction among firms can be ignored, thereby keeping the analysis relatively straightforward.

In line with other papers (see González and Jaumandreu, 1998; Sutton, 1991; and Levin and Reiss, 1988), it is assumed that the consumers' valuation of a given product depends on its quality, which can be improved through R&D expenditure. Consequently, the consumers' valuation is allowed to take the following functional form: $\Lambda_{it} = [s_{it}(x_{it-\tau})]^{\delta}$ in which s_{it} stands for product quality, $x_{it-\tau}$ denotes R&D expenditure and δ is the elasticity of the consumers' valuation with respect to quality. R&D investments affect product quality according to the relation $s_{it}(x_{it-\tau}) = x_{it-\tau}^{\phi}$ where $\phi < 1$ is the elasticity of quality with respect to R&D[11]. Notice that we are assuming that R&D does not immediately improve product quality but rather that it takes τ periods to become effective[12]. Quality is assumed to be constant at $s = 1$ if no R&D investments

---

[10] Then, the elasticity of demand with respect to price equals $-\sigma$.
[11] It is assumed that R&D is subject to diminishing returns to scale.
[12] Mansfield et al. (1971) report a median lag from R&D to innovation of about three years. Ravenscraft and Scherer (1982) cite survey responses from companies stating that 45% reported a typical time lag between the beginning of development and the first introduction of a new product of one to two years, 40% reported a lag between two and five years and 5% a lag of more than 5 years.

are made. Hence, firm i demand is $q_{it}(p_{it}, \Lambda_{it}) = z_{it} p_{it}^{-\sigma} x_{it-\tau}^{\phi\varepsilon}$ if it performs R&D and $q_{it}(p_{it}, \Lambda_{it}) = z_{it} p_{it}^{-\sigma}$ if it does not, where $\varepsilon = (\sigma - 1)\delta$ is the elasticity of demand with respect to quality.[13] Demand can also be expressed in a more compact way as follows: $q_{it}(p_{it}, \Lambda_{it}) = z_t p_{it}^{-\sigma}(1\{x_{it-\tau} = 0\} + 1\{x_{it-\tau} > 0\} x_{it-\tau})^{\phi\varepsilon}$.

3.2. Two-period problem without sunk costs

Before presenting the dynamic problem with sunk costs we first consider a simpler two-period setting that will serve to introduce many of the concepts we will use throughout. In this two-period setting firm *i* might choose to invest in R&D at *t*. If it does, then it reaps the benefits at $t + \tau$. Alternatively it might prefer not to invest in R&D. In such a case it gets standard non-R&D profits at $t + \tau$. We assume that for every euro spent on R&D each firm can expect to get a rebate $\rho_{it}^e \in [0,1]$ from the government. Hence, ρ_{it}^e is the expected share of subsidized R&D expenditure, something we shall later refer to as the subsidy coverage[14]. Also, let E_t be the expectations operator, parameter β stands for the discount factor, c_{it} represents marginal cost and f_{it} R&D fixed costs. Then, the expected gross operating profits of R&D performers are obtained by simultaneously choosing the price and the level of R&D expenditure that solve the following problem[15]:

$$\max_{p_{it+\tau}, x_{it}} E_t[\pi_{it}(p_{it+\tau}, x_{it})] = \beta^\tau E_t[(p_{it+\tau} - c_{it+\tau}) z_{it+\tau} p_{it+\tau}^{-\sigma} x_{it}^{\phi\varepsilon}] - (1 - \rho_{it}^e) x_{it} - f_{it}. \tag{3}$$

The first-order conditions lead to optimal price and R&D expenditure

$$p_{it+\tau}^*(c_{it+\tau}) = \frac{\sigma}{\sigma - 1} E_{it}[c_{it+\tau}] \tag{4}$$

[13] It seems sensible to assume that ε is below the unity. This assumption is met if $\delta \leq \frac{1}{(\sigma-1)}$.

[14] We model subsidies as a share of to-be-incurred R&D expenditures. This modeling is consistent with most subsidy granting schemes in Europe (see the 2006/C 323/01 issue of the Official Journal of the European Union for more details).

[15] Note that no equation has been specified for R&D capital stock formation. This is because we assume that only current R&D investments affect quality (or, what amounts to the same, that R&D capital fully depreciates from one period to the other). While not particularly realistic, this assumption ensures that the dynamic problem remains tractable. Marginal costs are assumed not to vary with the quantity produced.

$$x_{it}^{*}(\rho_{it}^{e}, c_{it+\tau}, z_{it+\tau}) = \left[\frac{\phi\varepsilon\beta^{\tau}E_{t}[A_{it+\tau}]}{(1-\rho_{it}^{e})}\right]^{1/1-\phi\varepsilon} \tag{5}$$

where $A_{it+\tau} = c_{it+\tau}^{1-\sigma}(\sigma-1)^{\sigma-1}\sigma^{-\sigma}z_{it+\tau}$. Plugging expressions (4) and (5) into the profit function gives rise to optimal current period profits, which turn out to be increasing in expected subsidies ρ_{it}^{e} and demand conditions $z_{it+\tau}$, and decreasing in marginal costs $c_{it+\tau}$ and fixed costs f_{it}:

$$E_{t}[\pi_{it}^{R\&D}(\rho_{it}^{e}, c_{it+\tau}z_{it+\tau}, f_{it})] = (1-\phi\varepsilon)\left(\frac{\phi\varepsilon}{(1-\rho_{it}^{e})}\right)^{\phi\varepsilon/1-\phi\varepsilon}\left(\beta^{\tau}E_{t}[A_{it+\tau}]\right)^{1/1-\phi\varepsilon} - f_{it}. \tag{6}$$

Proceeding analogously for the situation in which no R&D expenditures are incurred, it is immediate to obtain:

$$E_{t}[\pi_{it}^{NoR\&D}(c_{it+\tau}, z_{it+\tau})] = \beta^{\tau}E_{t}[A_{it+\tau}]. \tag{7}$$

The optimal participation rule is that the firm is R&D-active only if the profits generated by R&D are greater than the profits earned when not doing R&D. Because only equation (6) depends on subsidies, an optimal participation policy of this type can be characterized in terms of a threshold defined as the value of the subsidy for which the firm remains indifferent between performing R&D or not, that is, for which $(6)=(7)$:

$$\tilde{\rho}_{it}(c_{it+\tau}, z_{it+\tau}, f_{it}) = 1-\phi\varepsilon\left(\frac{1-\phi\varepsilon}{\beta^{\tau}E_{t}[A_{it+\tau}]+f_{it}}\right)^{1-\phi\varepsilon/\phi\varepsilon}\left(\beta^{\tau}E_{t}[A_{it+\tau}]\right)^{1/\phi\varepsilon}. \tag{8}$$

All firms with $\rho_{it}^{e} \geq \tilde{\rho}_{it}$ will self-select into R&D activities. Note that while ρ_{it}^{e} can only take values between 0 and 1 (as it is defined as the expected fraction of R&D expenditure covered by the subsidy), the threshold subsidy is fixed between minus infinity and one, $\tilde{\rho}_{it} \in (-\infty, 1]$, depending on the parameter values. Notice that the threshold subsidy is a negative function of $z_{it+\tau}$, ε and ϕ, while it is a positive function of $c_{it+\tau}$, f_{it}, σ and τ. Hence, firms with favourable demand shifters, high elasticity of quality with respect to R&D, high elasticity of demand with respect to quality, low marginal costs, a low elasticity of demand with respect to price (large market power) and short lags between R&D and profits should be less dependent on subsidies. Zero or negative thresholds denote that firms find it profitable to perform R&D no matter what their expected subsidies. By contrast, positive thresholds denote firms that rely on sufficiently large expected subsidies to engage in R&D. Given our assumptions and our modeling of subsidies as a share of to-be-made R&D expenditures all firms can be induced into R&D with a sufficiently large ρ_{it}^{e}. This is reasonable because even firms operating in very unfavourable conditions (with $\tilde{\rho}_{it} = 1$) will find it profitable to perform R&D if all expenditures are subsidized ($\rho_{it}^{e} = 1$).

Since R&D expenditure increases monotonically in the expected subsidies (see equation (5)), for any subsidy threshold $\tilde{\rho}_{it}$ there will exist a unique R&D threshold $\tilde{x}_{it} = x_{it}^{*}\ (\tilde{\rho}_{it})$. This implies that the optimal policy can be recast in terms of R&D expenditures. Plugging (8) into (5) we get the R&D threshold:

$$\tilde{x}_{it}(c_{it+\tau}, z_{it+\tau}, f_{it}) = \left(\frac{1}{1-\phi\varepsilon} + \frac{f_{it}}{(1-\phi\varepsilon)\beta^{\tau} E_t[A_{it+\tau}]} \right)^{1/\phi\varepsilon}. \tag{9}$$

The optimal decision is to perform R&D when $x_{it}^{*} \geq \tilde{x}_{it}$. Notably, ρ_{it}^{e} enters the optimal R&D equation (5) but not the R&D threshold (9). This will prove crucial for identification of the thresholds in the empirical exercise. We will expand on this issue later.

3.3. Dynamic setting: problem with sunk costs

Now, let us suppose that a sunk cost of K_{it} units is to be incurred every time a firm starts engaging in R&D.[16] In such a case it is clearly more costly to enter into R&D than it is to persist in R&D. In Baldwin's (1989) words, *sunk costs imply that it is easier for firms to stay "in" than it is to get "in"*. This circumstance can favour cases in which firms find it optimal to persist in R&D even when profit levels are lower than those that

[16] Sunk costs are to be incurred if a firm performs R&D for the first time but also if a firm stopped performing R&D for one period. In other words we assume that you cannot keep your R&D facilities idle.

could be obtained by abandoning innovative activities, since by doing so firms avoid future re-entry costs (Clerides et al., 1998). Thus, firms face a dynamic optimization problem in which they must decide, in each period, whether to perform R&D or not on the grounds of their expectations over ρ^e, c, z and f. Therefore, the firm will plan its participation in R&D activities in order to maximize its present discounted profits (since our interest lies on subsidies in what follows we abstract from c, z and f and simplify notation by writing $\pi_{it}^{R\&D}(\rho_{it}^e) = E_t[\pi_{it}^{R\&D}(\rho_{it}^e, c_{it+\tau} z_{it+\tau}, f_{it})]$ and $\pi_{it}^{NoR\&D} = E_t[\pi_{it}^{NoR\&D}(c_{it+\tau}, z_{it+\tau})]$):

$$V_{it} = \max_{\{y_{it+s}\}_0^{\infty}} E_{it} \sum_{s=0}^{\infty} \beta^s \left[y_{it+s} \left(\pi_{it+s}^{R\&D}(\rho_{it+s}^e) - (1 - y_{it+s-1}) K_{it} \right) + (1 - y_{it+s}) \pi_{it+s}^{NoR\&D} \right] \qquad (10)$$

where y_{it} is a binary variable with value one if the firm performs R&D at period t and value zero otherwise.[17] It amounts to the same thing, and at the same time it is much simpler, to characterize the optimal participation policy by choosing the y_t that satisfies the Bellman equation corresponding to the above expression:

$$V_{it} = \max_{y_{it}} \left[\mathrm{y}_{\mathrm{it}} \left(\pi_{it}^{R\&D}(\rho_{it}^e) - (1 - y_{it-1}) K_{it} \right) + (1 - y_{it}) \pi_{it}^{NoR\&D} + \beta E_{it} \left(V_{it+1} \mid y_{it} \right) \right] \qquad (11)$$

The profit-maximizing firm will calculate the value function for both $y_{it} = 1$ and $y_{it} = 0$ and will choose the option yielding the highest value. In this kind of infinite horizon problem with entry costs it is well known that the optimal participation strategy can be characterized in terms of two threshold values defined as the realization of expected

[17] In such a context besides subsidies for engaging in R&D we might also consider subsidies aimed at lowering the sunk costs of entry. Such subsidies could be introduced within equation (10) as a lump sum quantity to be subtracted from K_{it}. However, there are two reasons why this might not be such a good idea. Firstly, sunk costs are difficult to calculate due to their somewhat tenuous nature and agencies are reluctant to subsidize quantities that cannot be directly observed. They rather prefer to subsidize a percentage of a firm's ordinary R&D expenditure. Secondly, most of the datasets containing information on subsidies do not specify what the subsidies are for and simply provide an overall amount. Hence, it is impossible for the researcher to identify the exact nature of the subsidy and to determine whether it is aimed at lowering entry costs or not.

subsidies for which the firm is indifferent to being active and inactive[18]. This is due to the fact that the indifference condition depends on whether firms have previous experience in R&D. The indifference condition is given by:

$$\pi_{it}^{R\&D}(\tilde{\rho}_{it}) - \pi_{it}^{NoR\&D} + \beta\psi_{it+1}\left[\tilde{\rho}_{it}\right] = (1 - y_{it-1})K_{it} \tag{12}$$

where $\psi_{it+1}\left[\tilde{\rho}_{it}\right] = \left[E_t(V_{it+1} \mid \tilde{\rho}_{it}, y_{it} = 1) - E_t(V_{it+1} \mid \tilde{\rho}_{it}, y_{it} = 0)\right]$ is the discounted expected value of the advantage that can be enjoyed at period $t+1$ by a firm that is already R&D-performing at period t. Baldwin (1989) refers to this advantage as an incumbency premium. Note that while the thresholds are implicitly defined by equation (12), there is no analytical expression for them. Nevertheless, provided that certain conditions hold, the period t optimal entry-exit strategy can be depicted as in Figure 2. We will refer to the threshold values as $\tilde{\rho}_{it}^{E}$ when $y_{it-1} = 0$ and $\tilde{\rho}_{it}^{C}$ when $y_{it-1} = 1$ with $\tilde{\rho}_{it}^{E} \geq \tilde{\rho}_{it}^{C}$. The superscripts E and C have been chosen to reflect the fact that one threshold characterizes "Entry" while the other characterizes "Continuation" of R&D. Accordingly, firm's optimal entry-exit strategy will be to perform R&D only if $\rho_{it} \geq \tilde{\rho}_{it}^{E}$ when $y_{it-1} = 0$ or if $\rho_{it} \geq \tilde{\rho}_{it}^{C}$ when $y_{it-1} = 1$.

[INSERT FIGURE 2]

As shown in the two-period setting without sunk costs, firms' optimal policy can likewise be stated in terms of optimal R&D expenditure. Since R&D expenditure increases monotonically in the expected subsidies (see equation (5)), for any pair of subsidy thresholds $\tilde{\rho}_{it}^{E}$ and $\tilde{\rho}_{it}^{C}$ there will exist a unique pair of R&D thresholds $\tilde{x}_{it}^{E} = x^{*}(\tilde{\rho}_{it}^{E})$ and $\tilde{x}_{it}^{C} = x^{*}(\tilde{\rho}_{it}^{C})$. Thus, the optimal decision is to perform R&D when $x_{it}^{*} > \tilde{x}_{it}^{E}$ and $y_{it-1} = 0$ or when $x_{it}^{*} > \tilde{x}_{it}^{C}$ and $y_{it-1} = 1$, and to refrain from R&D otherwise.

4. Econometric modeling

[18] Actually, the optimal participation strategy can be defined in terms of either of the state variables (ρ^{e}, c, z and f in this case), but for our purposes it is convenient to define them as a function of ρ^{e}.

Econometrically, firms' optimal participation policy can be cast in terms of a type-2 tobit specification in which R&D expenditure x_{it}^{*} is observed only when $x_{it}^{*} - \tilde{x}_{it}^{E} > 0$ for first-time R&D performers and when $x_{it}^{*} - \tilde{x}_{it}^{C} > 0$ for continuing R&D performers. Assuming that the logs of x_{it}^{*} and $\tilde{x}_{it}^{E}$ or $\tilde{x}_{it}^{C}$ can be linearly approximated by a set of reduced form determinants, the tobit model is defined by the following equations[19]:

$$\ln x_{it}^{*} = \gamma sub_{it} + w_{1it}\beta_1 + \alpha_{1i} + \varepsilon_{1it} \tag{13}$$

$$\ln \tilde{x}_{it} = -\eta y_{it-1} + w_{0it}\beta_0 + \alpha_{0i} + \varepsilon_{0it} \tag{14}$$

where $\tilde{x}_{it} = \tilde{x}_{it}^{E}$ when $y_{t-1} = 0$ and $\tilde{x}_{it} = \tilde{x}_{it}^{C}$ when $y_{t-1} = 1$. As for the optimal R&D equation (equation (13)), $sub_{it} = -\ln(1-\rho_{it}^{e})$, which implies that expected subsidies are expressed in the way they appear in equation (5). The remaining determinants of optimal R&D, namely the elasticities ε, ϕ and σ, the marginal costs c, and the demand shifters z, are unobservable and need to be approximated by a set of exogenous or predetermined variables w_{1it} (this will be explained in section 5.1). Similarly, the thresholds are assumed to be a function of the same variables contained in w_{1it} plus a number of other variables that account for fixed costs f in such a way that w_{0it} contains at least all the variables that appear in w_{1it}. In addition, as suggested by the analytical framework, we suspect that the threshold might take two different values depending on a firm's past R&D. For this reason, we allow it to be a function of y_{it-1}, a dummy variable that takes value one if the firm performed R&D at $t-1$ and zero otherwise. In this way, the continuation threshold is lower than the entry threshold by η , a parameter to be estimated. We assume that the two thresholds differ only by the parameter η. By examining the significance and the magnitude of η it is possible to conclude whether there are two thresholds rather than one and to measure the distance between them. Finally, both the optimal R&D and the threshold equations include time-invariant individual effects, α_{1i} and α_{0i}, and idiosyncratic error terms, ε_{1it} and ε_{0it}.

[19] Taking logarithms is a necessary step if we are to assume normality given that R&D expenditures follow a lognormal distribution.

4.1. Identification of the thresholds with a dynamic panel data type-2 tobit model

Clearly, the thresholds are not observable in practice, which implies that the parameters of equation (14) cannot be estimated directly. Fortunately, we can observe a firm's decision to perform R&D, which contains information about the relationship between optimal and threshold R&D. Specifically, R&D performance takes place when $x_{it}^{*} - \tilde{x}_{it}^{E} > 0$ for new R&D performers and when $x_{it}^{*} - \tilde{x}_{it}^{C} > 0$ for ongoing R&D performers. More formally, this can be expressed in the classical type-2 tobit formulation with the following selection and level equations:

$$y_{it} = 1[\eta y_{it-1} + \gamma sub_{it} + w_{0it}\beta_2 + \alpha_{2i} + \varepsilon_{2it} > 0] \tag{15}$$

$$y_{1it} = \begin{cases} \ln x_{it}^{*} & if \;\; y_{it} = 1 \\ 0 & if \;\; y_{it} = 0 \end{cases} \tag{16}$$

where $\beta_2 = \beta_1 - \beta_0$, $\alpha_{2i} = \alpha_{1i} - \alpha_{0i}$, $\varepsilon_{2it} = \varepsilon_{1it} - \varepsilon_{0it}$ and x_{it}^{*} is given by equation (13)[20]. Under certain conditions, in a maximum likelihood estimation framework, the parameters of the threshold equation (η and β_0) can be recovered through the relationship between the parameters of the selection and the level equations. As discussed in Nelson (1977) an exclusion restriction, in our case the absence on theoretical grounds of the subsidy variable in the threshold equation, is a sufficient condition for the identification of all parameters of the model.

4.2. The relationship between true state dependence and the thresholds

The main feature of selection equation (15) is that it includes the lag of the dependent variable among the set of regressors. Algebraically, this is a very obvious derivation of the fact that the threshold equation includes dynamics. Conceptually, however, the mechanism by which the existence of the two thresholds results in a dynamic selection equation is very interesting and merits careful consideration.

Dynamic selection equations enable us to identify whether R&D performance exhibits persistence, and whether this persistence is attributable to true state dependence as

[20] For variables like fixed costs that appear in (14) but not in (13) the corresponding coefficient of β_2 equals $-\beta_0$.

opposed to spurious state dependence. True state dependence implies that a causal behavioural effect exists in the sense that the decision to undertake R&D in one period enhances the probability of R&D being undertaken in the subsequent period. In the presence of sunk costs two thresholds must exist if true state dependence is prevalent. To understand why, note that for any optimal R&D that lies between the entry and the continuation threshold, $\tilde{x}_{it}^{C} < x_{it}^{*} < \tilde{x}_{it}^{E}$, present R&D performance occurs thanks to the past performance of R&D. The wider the gap between the two thresholds, i.e. the higher the sunk costs, the higher is the chance of having true state dependence.

Figure 3 illustrates the relationship between the thresholds and true state dependence. It considers two optimal R&D paths that take different values in the initial period but are identical thereafter. The deviation in the initial period is not trivial, though, and leads to different R&D decisions: path 1 entails R&D performance at $t=0$ while path 2 does not. This initial departure allows us to evaluate, for periods $t=1$ to $t=4$, the relevance of previous experience in explaining present R&D performance. This evaluation is conducted for three different scenarios that consider varying distances between the thresholds reflecting the magnitude of R&D sunk costs. As the continuation threshold gradually approaches that of entry and the gap between the thresholds shrinks, the importance of past experience in accounting for present R&D performance decreases and true state dependence vanishes. For instance, in case 1 where the distance between the thresholds is substantial, experience is found to have considerable impact: path 1 leads to R&D performance from $t=1$ onwards while path 2 never results in R&D performance. The effect of previous experience declines in case 2, where the distance between the thresholds is smaller. Here, previous experience only explains R&D performance at $t=1$. Finally, when there is a single threshold, as in case 3, previous experience is irrelevant for explaining R&D performance. In the estimation framework of equation (15), case 1 should lead to significant and sizeable estimates of η while case 2 should lead to significant but modest estimates and case 3 to values insignificantly different from zero.

[INSERT FIGURE 3]

4.3. Maximum likelihood estimation

The estimation of dynamic panel data sample selection models poses two main problems: the treatment of unobserved individual effects and the so-called problem of the initial conditions. The modeling of the former through fixed effects leads to the "incidental parameters" problem, which results in inconsistent maximum likelihood estimators when the number of periods is small (Neyman and Scott, 1948). The latter arises because of the fact that, for variables generated by stochastic dynamic processes, the first observation (that which initializes the process) is correlated both with future realizations of the variable (due to state dependence) and with the unobservable individual term (given that the unobservable term is part of the process that generates the variable). Consequently, unless the first observation in the process (i.e., the initial condition) is accounted for, the lagged dependent variable will be correlated with the unobservable term and the estimates will be inconsistent[21].

We use the method proposed by Raymond et al. (2010) which provides simple, satisfactory solutions to both of these problems: in light of the shortcomings of the fixed effects approach, they assume the individual effects α_{1i} and α_{2i} to follow a joint distribution. Moreover, they adopt Wooldridge's (2005) solution to the initial conditions problem, which involves modeling the individual term as a linear function in the explanatory variables and the initial conditions

$$\alpha_{1i} = \alpha_1^0 + \alpha_1^1 y_{i0} + \overline{w}_{1i}\alpha_1^2 + a_{1i} \tag{17}$$

$$\alpha_{2i} = \alpha_2^0 + \alpha_2^1 y_{i0} + \overline{w}_{0i}\alpha_2^2 + a_{2i} \tag{18}$$

where α_1^0 and α_2^0 are constants, $\overline{w}_{1i}$ and $\overline{w}_{0i}$ are the Mundlak within-means (1978) of the explanatory variables and y_{i0} is the initial condition, which takes a value of one if the firm performs R&D in the first year of the sample used for conducting the estimates and 0 otherwise. The vectors $(\varepsilon_{1it}, \varepsilon_{2it})'$ and $(a_{1i}, a_{2i})'$ are assumed to be independently and identically (over time and across individuals) normally distributed with means zero and covariance matrices:

[21] Heckman (1981) provides a good account of the problem of initial conditions.

$$\Omega_{\varepsilon1\varepsilon2} = \begin{pmatrix} \sigma_{\varepsilon1}^2 & \rho_{\varepsilon1\varepsilon2}\sigma_{\varepsilon1}\sigma_{\varepsilon2} \\ \rho_{\varepsilon1\varepsilon2}\sigma_{\varepsilon1}\sigma_{\varepsilon2} & \sigma_{\varepsilon2}^2 \end{pmatrix} \text{ and } \Omega_{a1a2} = \begin{pmatrix} \sigma_{a1}^2 & \rho_{a1a2}\sigma_{a1}\sigma_{a2} \\ \rho_{a1a2}\sigma_{a1}\sigma_{a2} & \sigma_{a2}^2 \end{pmatrix}$$

With the above assumptions, the likelihood function of one individual, starting from *t=1* and conditional on the means of the regressors and the initial conditions, is written as

$$L_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{t=1}^{T} L_{it}\left(y_{1it} \mid y_{i0}, y_{it-1}, \overline{w}_i, a_{1i}, a_{2i}\right) g\left(a_{1i}, a_{2i}\right) da_{1i} da_{2i} \tag{19}$$

where $\prod_{t=1}^{T} L_{it}\left(y_{1it} \mid y_{i0}, y_{it-1}, \overline{w}_{it}, a_{1i}, a_{2i}\right)$ denotes the likelihood function once the individual effects have been integrated out and can be treated as fixed, and $g\left(a_{1i}, a_{2i}\right)$ stands for the bivariate normal density function of $(a_{1i}, a_{2i})'$. The double integral in equation (19) will be approximated by a "two-step" Gauss-Hermite quadrature (see Raymond et al. (2010) for a derivation of the "two-step" Gauss-Hermite quadrature expression). Next, treating the individual effects as fixed and using the standard properties of the bivariate normal distribution, the partial conditional likelihood function for firm *i* at period *t* can be written as follows

$$\begin{aligned} L_{it} = \Phi\left[-\left(\delta_0 sub_{it} + A_{it} + \widetilde{a}_{2i}\right)\right]^{(1-y_{it})} \times \Bigg[\frac{1}{\sigma_{\varepsilon1}} \phi\left(\frac{y_{1it} - \gamma sub_{it} - B_{it} - a_{1i}}{\sigma_{\varepsilon1}}\right) \\ \times \Phi\left(\frac{\delta_0 sub_{it} + A_{it} + \widetilde{a}_{2i} + \rho_{\varepsilon1\varepsilon2}\left(y_{1it} - \gamma sub_{it} - B_{it} - a_{1i}\right)/\sigma_{\varepsilon1}}{\sqrt{1-\rho_{\varepsilon1\varepsilon2}^2}}\right)\Bigg]^{y_{it}} \end{aligned} \tag{20}$$

where

$$\delta_0 = \frac{\gamma}{\sigma_{\varepsilon2}} \tag{21}$$

$$A_{it} = [\eta y_{it-1} + w_{0it}\beta_2 + \alpha_2^0 + \alpha_2^1 y_{i0} + \overline{w}_i\alpha_2^2]/\sigma_{\varepsilon2} \tag{22}$$

$$\widetilde{a}_{2i} = \frac{a_{2i}}{\sigma_{\varepsilon2}} \tag{23}$$

$$B_{it} = w_{1it}\beta_1 + \alpha_1^0 + \alpha_1^1 y_{i0} + \overline{w}_{1i}\alpha_1^2. \tag{24}$$

The presence of γ in both the optimal R&D equation [equation (13)] and the selection equation [equation (15)] together with the exclusion of the subsidy coverage from the threshold equation [equation (14)], allows us to identify the standard error of ε_{2it}: $\sigma_{\varepsilon 2} = \gamma / \delta_0$. Knowing $\sigma_{\varepsilon 2}$, it is possible to recover all the parameters of the threshold equation via (22). So correct identification of the parameters of the threshold equation critically depends on parameters δ_0 and γ. We will in turn discuss how to correctly identify these two parameters.

4.4. Identification of γ and δ_0.

Public agencies may be encouraged to support projects with the best technical merits and the highest potential for commercial success. As these projects typically have high private returns they are likely to be undertaken even in the absence of the support. In other words, subsidies are granted by agencies according to the contemporary effort and performance of firms, and hence are presumably endogenous. This implies that the compound error terms of the levels and the selection equations $a_{1i} + \varepsilon_{1it}$ and $a_{2i} + \varepsilon_{2it}$ are likely to be positively correlated with ρ_{it}^e leading to upward biased estimates of γ and δ_0.

To solve this problem we assume that the subsidies to which firms react are expected in advance, along the lines of González et al. (2005). However, we construct a slightly different measure of expected subsidies drawing on the information we have on subsidy applicants. González et al. (2005) calculate the expected subsidy coverage as follows:

$$\rho_{it}^e = E(\rho_{it} \mid z_{it}^{\rho}) = P(\rho_{it} > 0 \mid z_{it}^{\rho}) E(\rho_{it} \mid z_{it}^{\rho}, \rho_{it} > 0) \qquad (25)$$

where $P(\rho_{it} > 0 \mid z_{it}^{\rho})$ is the probability of receiving a subsidy (joint probability of applying for a subsidy and receiving the subsidy) and $E(\rho_{it} \mid z_{it}^{p}, \rho_{it} > 0)$ is the expected value of the subsidy for successful applicants.

Unlike in González et al. (2005) our expected subsidy coverage is not positive for all firms but just for subsidy applicants. This gives more within variation to the expected subsidy shares and enables us to control for fixed effects via the inclusion of Mundlak means, which will ultimately result in a better identification of γ and δ_0. We calculate the expected subsidy coverage as follows:

$$\rho_{it}^{e} = \begin{cases} P(\rho_{it} > 0 \mid z_{it}^{\rho}, ap_{it} = 1)E(\rho_{it} \mid z_{it}^{\rho}, \rho_{it} > 0) & if\ ap_{it} = 1 \\ 0 & if\ ap_{it} = 0 \end{cases} \tag{26}$$

where ap_{it} is a dummy with value one if firm *i* applies for a subsidy in year *t*, $P(\rho_{it} > 0 \mid z_{it}^{\rho}, ap_{it} = 1)$ is now the probability of receiving a subsidy among applicants and $E(\rho_{it} \mid z_{it}^{p}, \rho_{it} > 0)$ is the expected value of the subsidy for successful applicants. We estimate $P(\rho_{it} > 0 \mid z_{it}^{\rho}, ap_{it} = 1)$ by means of a probit with parameters λ_1 and assume $\ln(\rho_{it} \mid z_{it}^{\rho}, \rho_{it} > 0) \sim N(z_{it}^{\rho}\lambda, \sigma^2)$ to estimate $E(\rho_{it} \mid z_{it}^{p}, \rho_{it} > 0)$ by means of an OLS regression with parameters λ_2. We use an augmented version of the specification proposed in González et al. (2005) to estimate the parameters λ_1 and λ_2 used to construct ρ_{it}^{e} (see appendix A). We assume that the expected coverage ratio is uncorrelated with the idiosyncratic and the individual-specific error terms in (13) and (15).

5. Empirical specification and results

5.1. Empirical specification

Optimal R&D equation – The dependent variable used in the main equation of interest is the natural logarithm of R&D expenditure. The explanatory variable of interest is $-\ln(1-\hat{\rho}_{it}^{e})$. The main control is the Mundlak mean of $-\ln(1-\hat{\rho}_{it}^{e})$. The remaining explanatory variables are derived from equation (5). Some of these are lagged by one period to ensure that they are predetermined. Average variable costs (lagged by one period) are used as a proxy for future marginal costs ($c_{it+\tau}$). Future demand shifters (

$z_{it+\tau}$) are captured by two dummy variables (both lagged by one period) that report whether the main market of the firm is in recession or expansion. The elasticity of demand with respect to quality (ε) and of product quality with respect to R&D (ϕ) are approximated by the advertising/sales ratio (lagged) and the average industry patents. Finally, a firm's market share and a dummy variable representing concentrated markets (both lagged by one period) are used as indicators of the elasticity of demand with respect to price (σ).

Selection equation –The dependent variable of the selection equation is a dummy indicating whether or not the firm performed R&D at period t. The explanatory variables in the selection equation are a combination of the variables in the levels and the threshold equations (see equation (15)). Thus, apart from the variables included in the optimal R&D equation the selection equation also contains some extra variables specific to the threshold equation. These variables are the lagged dependent variable (y_{it-1}) and a set of variables aimed at capturing fixed costs (f_{it}): the presence of foreign capital, quality controls and the employment of highly skilled workers. We also control for the subsidy applicant dummy (ap_{it}).

It is reasonable to assume that larger firms will make larger R&D investments. For this reason, in addition to all the variables listed above, we include a set of employment-size dummies and the total sales (in logarithms) of the firm in both equations. Sales are assumed to be predetermined given that they are only affected by year $t-\tau$ R&D expenditures. Notice that including sales in the right hand side is equivalent to adopting a Dorfman and Steiner (1954)-type expression (see González et al., 2005). Starting from equation (5) of the underlying theoretical model and assuming $\tau = 0$ we get that R&D effort is given by $\ln(x_{it}^{*} / p_{it}^{*} q_{it}^{*}) = -\ln(1-\rho_{it}^{e}) + \ln(\phi\varepsilon) - \ln(\sigma)$ or, what is the same, that optimal R&D expenditures are given by $\ln(x_{it}^{*}) = -\ln(1-\rho_{it}^{e}) + \ln(\phi\varepsilon) - \ln(\sigma) - \ln(p_{it}^{*} q_{it}^{*})$. This implies that, given our specification, the parameter (γ) in the optimal R&D equation should be one. For $\tau > 0$ we cannot get to such a compact expression, but if we are willing to assume that the departure from the situation in which $\tau = 0$ is not too sharp, then the parameter should still be close to one (even though not necessarily one).

Moreover, we also include year and industry dummies to account for variations in the business cycle and any sector-specific characteristics. The explanatory variables (other than $-\ln(1-\hat{\rho}_{it}^{e})$) have little within variation and are highly correlated with their Mundlak means. After experimenting with different specifications we resolved not to include the Mundlak means of the control variables in the regressions[22]. Descriptive statistics and definitions of all the explanatory variables are reported in Tables 5 and 6.

[INSERT TABLE 5]

[INSERT TABLE 6]

5.2. Estimation results

Table 7 shows the estimates obtained with the dynamic panel data type-2 tobit model. In the optimal R&D equation the parameter associated with the expected subsidy coverage is substantially below unity. This estimate would suggest partial crowding-out. If the subsidy coverage increases by 1 percentage point, hence the R&D costs supported by the firm decrease by 1 percent, R&D increases by only 0.3 percent. This result is quite different from the point estimates of González et al. (2005) and Takalo et al. (2011) who get a coefficient close to one. One potential explanation for this low coefficient is that in our regressions we consider some firms with subsidies covering almost 100% of their R&D expenditures. These large subsidy shares might not fit our modeling of subsidies as a share of to-be-incurred R&D expenditures and might well be aimed at financing endeavours other than R&D. Indeed, CDTI's (Spain's national agency of technology) upper bound for the share of covered R&D costs was 60% until 2007 and increased up to 75% in 2007. This would explain the lack of sensitivity of R&D with respect to the subsidy coverage. When we restrict to firms with subsidies lower than 60% or 75% the coefficient comes closer to one (see Table 8) meaning that subsidies are not misused (every euro of subsidy is invested in R&D)[23].

---

[22] This implies that the coefficients of the explanatory variables will be the sum of their direct effects and their correlations with the individual effects, so that they should be interpreted as plain correlations rather than as causal effects. This is not problematic in our setting given that we mainly need the controls to establish the height of the thresholds. The results we will present in the next sections are robust to the inclusion of the Mundlak means in the regressions.

[23] The same happens when we use the actual subsidy coverage instead of the expected subsidy coverage. The coefficient of the levels equation is insignificant when we use all subsidized firms, and it becomes significant and close to one when we restrict the sample to firms with subsidy shares below 60%. Notice that it is possible to estimate the optimal R&D equation with the actual subsidy coverage but not the selection equation because all subsidized firms carry out R&D. This causes actual subsidies to perfectly

[INSERT TABLE 7]

[INSERT TABLE 8]

The expected subsidy coverage parameter is also significant in the selection equation, indicating that subsidies not only affect the level of investment in R&D but also the decision to perform R&D (the point estimate does not vary when we restrict the sample to firms with 60% and 75% subsidy coverages). An illustrative magnitude of subsidies' inducement effects is given by the average marginal increase in firms' probability of performing R&D caused by a discrete change in the expected subsidy coverage from zero to a positive magnitude (assuming $y_{it-1} = 0$ and fixing all other regressors at their mean). A discrete change in the expected subsidy coverage from 0 to 30% (0 to 60%) leads to an increase in firms' probability of performing R&D of 17 (55) percentage points.

The significance of the lagged dependent variable in the selection equation indicates that true state dependence exists. This result is in line with the findings of Peters (2009) and Mañez et al. (2009). We can conclude that there is a behavioural effect: firms that perform R&D in a given period have a 37% higher probability of performing R&D in the next period than firms that did not perform R&D (see the average partial effect reported at the bottom of Table 8). A direct consequence of the existence of true state dependence is that the R&D threshold also depends on past R&D performance giving rise to an entry and a continuation threshold. The distance between the two thresholds (in logarithms) is $\eta = 0.2$ (see equation (14)), meaning that the continuation threshold is 20% lower than the entry threshold[24].

---

predict R&D performance. So it is not possible to estimate the subsidies inducement effects with the actual subsidy coverage alone.

[24] This distance between the two thresholds should be seen as a lower bound given that the coefficient of the expected subsidy coverage in the levels equation is possibly greater than our point estimate $\hat{\gamma} = 0.31$. If we use the point estimate obtained in column (2) of Table 8 ($\hat{\gamma} = 0.64$) the distance between the thresholds doubles: $\eta = (0.64 / 2.48) * 1.60 = 0.41$ (recall that the coefficients of the threshold equation are obtained as $b_0 = b_1 - \sigma_{\varepsilon 2} b_2$ where b_1 and b_1 are the coefficients of the optimal R&D and the selection equations respectively and $\sigma_{\varepsilon 2} = \gamma / \delta_0$).

The signs of the coefficients of the other explanatory variables are largely in agreement with the results reported in González et al. (2005) and the predictions from the analytical section. The advertising to sales ratio, as a proxy of the elasticity of demand with respect to quality ε, has a positive and significant impact on both R&D expenditure and on the decision to perform R&D but is not significant in the threshold. High average variable costs (as a proxy for marginal cost c) seem to be an obstacle for R&D performance but do not significantly affect the thresholds. The quality controls and skilled labour dummies, designed to capture fixed costs (excluded from the optimal R&D equation on theoretical grounds) are found to have a positive and significant effect on R&D performance and a negative effect on the thresholds (although this negative effect is only significant in the case of the skilled labour dummy). Finally, the variables aimed at accounting for scale effects, such as the set of size dummies and the sales volume, have a positive and significant impact on optimal R&D expenditure. The sales volume also positively affects the propensity to perform R&D and the threshold. This is a logical result that confirms that larger firms make larger R&D investments reflecting their larger capacity or the more pressing requirement to achieve a perceptible impact in their already large volume of business.

5.3. Robustness checks

The estimated subsidy coefficients in the optimal R&D and the selection equations do not change much when we estimate the optimal R&D equation without the selection equation, controlling or not for individual effects, assuming them to be fixed or random, and controlling or not for the other control variables (these results are not reported but are available upon request). We also experimented with two alternative measures of the expected subsidy coverage. The first one ($\hat{\rho}_{it}^{e}\_1$) is calculated via expression (2A) in appendix A, using the same static specification as in González et al. (2005) but predicting the subsidy coverage on the basis of the probability of being successful in the subsidy applications for applicants only, assigning a value zero to non-applicants. The second one ($\hat{\rho}_{it}^{e}\_2$) is calculated via expression (3A) in appendix A and is exactly equivalent to the one in González et al. (2005), i.e. with the same specification and predicted with the probability of getting a subsidy for all firms in the sample.

Table 9 (column (1)) shows that our results still hold when we use $\hat{\rho}^e_{it}\_1$ instead of $\hat{\rho}^e_{it}$. It also shows that $\hat{\rho}^e_{it}\_2$ (the expected subsidy coverage measured as in González et al. (2005)) is not resistant to the inclusion of the Mundlak means. Column (2) reports static estimates equivalent to the ones of González et al. (2005). The results are very similar to theirs (even though we use different survey years). When we include the lagged R&D dummy in column (3) the coefficient of the selection equation is still significant but much lower. This suggests that González et al. (2005) attribute to subsidies an effect that may be due to persistence. When we include the Mundlak mean in column (4), subsidies are not significant anymore because $\hat{\rho}^e_{it}\_2$ does not have enough within variation to disentangle its direct effect from the Mundlak means.

[INSERT TABLE 9]

6. Policy implications

6.1. R&D and subsidy coverage thresholds

First, we would like to characterize the distributions of the entry and continuation R&D thresholds. They are calculated from the estimated parameters of equation (14) according to the following expressions:

$$\widetilde{x}^E_{it} = \exp(w_{0it}\hat{\beta}_0 + (1/2)\hat{\sigma}^2_{\varepsilon 0}) \tag{25}$$

$$\widetilde{x}^C_{it} = \exp(-\hat{\eta} + w_{0it}\hat{\beta}_0 + (1/2)\hat{\sigma}^2_{\varepsilon 0}) \tag{26}$$

Similarly, the subsidy thresholds are calculated using the estimated parameters of equations (13) and (14) as the subsidies that make the firms indifferent between performing R&D or not, i.e., equalizing optimal and threshold R&D (equations (13) and (14)), giving the following expressions[25]:

[25] Recall that threshold R&D is the level of R&D expenditure that makes the firm indifferent to performing R&D or not.

$$\tilde{\rho}_{it}^{E} = 1 - \exp\left(\frac{w_{1it}\hat{\beta}_1 - w_{0it}\hat{\beta}_0}{\hat{\gamma}}\right) \quad (27)$$

$$\tilde{\rho}_{it}^{C} = 1 - \exp\left(\frac{w_{1it}\hat{\beta}_1 + \hat{\eta} - w_{0it}\hat{\beta}_0}{\hat{\gamma}}\right). \quad (28)$$

Kernel densities of the R&D and subsidy thresholds are provided in Figure 4, which is complemented by Table 10. [26] As expected, continuation thresholds take on average lower values than those adopted by entry thresholds. For instance, while only 42% of the entry R&D thresholds are below 50,000€, as much as 47% of the continuation R&D thresholds are below 50,000€. Focusing on the largest values, 27% of the entry thresholds are above 200,000€ while 23% of continuation thresholds reach this value. Regarding subsidy coverage, around 44% of the entry thresholds concentrate in values higher than 60% which implies that most firms need to have their R&D expenditure almost entirely subsidized in order for them to engage in R&D. It is equally true that 18% of the entry thresholds take negative values meaning that there is a mass of firms which does not require subsidies to engage in R&D. Not surprisingly, the percentage of firms with negative continuation thresholds is much larger, with a value close to 38%. This implies that almost half of the firms in the sample are self-sufficient to continue performing R&D in the absence of public support.

[INSERT FIGURE 4]

[INSERT TABLE 10]

6.2. Permanent inducement effects

The second question we set out to address is whether subsidies can achieve permanent inducement effects. By knowing the entry and continuation subsidy thresholds, it is possible to classify firms in three different scenarios according to their dependence on subsidies. The first scenario considers firms that have positive entry and exit thresholds ($\tilde{\rho}^{E} > 0 \;\&\; \tilde{\rho}^{C} > 0$). These firms should be permanently subsidized to ensure the

[26] Figure 4 only shows the range (0, 600000) for threshold R&D which is where most of the observations concentrate (90% and 93% of entry and continuation R&D thresholds lie in this interval). Similarly, it only shows the range (-1, 1) for threshold subsidies (99% and 84% of the entry and the continuation subsidy coverage thresholds lie in this interval). Note, however, that the kernel densities have been calculated using all the observations in the sample.

profitability of their R&D activities. The second scenario ($\tilde{\rho}^E \leq 0$ & $\tilde{\rho}^C \leq 0$) considers firms that have negative entry and exit thresholds and, hence, find R&D profitable even in the absence of subsidies. The third scenario ($\tilde{\rho}^E > 0$ & $\tilde{\rho}^C \leq 0$), considers firms with positive entry thresholds but negative continuation thresholds. The last scenario opens up the possibility of using subsidies to induce permanent entry into R&D through temporary increases in a firm's expected subsidies.

Column 1A in Table 11 shows that 20% of the observations in the sample require subsidies to start performing R&D but can continue performing R&D without them; 18% can perform R&D regardless of the subsidies; and the remaining 62% always require a subsidy to persist in R&D activities. Interestingly (see column 1B), 60% of the observations that only need entry subsidies are actually already performing R&D, while the other 40% has still to be induced into engaging in R&D activities. Further, almost all the firms (93%) that do not depend on subsidies are R&D performers and virtually none (just 5%) of the firms that always need subsidies perform R&D.

[INSERT TABLE 11]

In column (2) we refer to firms instead of observations and find that some of the firms that need only a trigger subsidy to become stable R&D performers change from one scenario to another over the sample years. For instance, 11% of the firms need entry subsidies in certain periods but can enter into R&D without such a requirement in others. Another similarly sized group (8%) alternates between periods of dependence on entry subsidies and periods of dependence on both entry and continuation subsidies.

In column (3) we report the values for the whole population of Spanish manufacturing firms[27]. The figures show that 25% (9+9+7) of Spanish manufacturing firms need subsidies to enter into R&D but not to continue. Only 5% of the firms can perform R&D without subsidies (almost all of which actually perform R&D). This value is notably lower than that obtained in column (2), reflecting the fact that this group is

[27] We are able to undertake this exercise because the ESEE has a known representativeness. The number of small (between 10 and 200 employees) and large (more than 200 employees) firms included in the sample amounts to 5% and 50% of the whole population respectively. Hence, all we need to do in order to build representative proportions is to multiply, where appropriate, the number of small and large firms by 20 and 2 respectively.

comprised mainly of large firms, which are in fact over-represented in the ESEE sample. The opposite is true for the proportion of firms that need subsidies to both start and continue performing R&D which amounts to 70% of the population.

On the basis of these results, we can conclude that there is a case for using subsidies to induce firms to go permanently into R&D by means of one-shot trigger subsidies. Around 10.7% (9%*(1-0.56) + 9%*(1-0.72) + 7%*(1-0.4)) of Spanish manufacturing firms can be permanently brought into R&D by means of trigger subsidies (this number lowers to 6.5% (9%*(1-0.56) + 9%*(1-0.72)) if we disregard firms in row 5 which require continuation subsidies in some periods).

Table 12 provides information on the distribution of entry subsidies of those firms that can be permanently induced. Remarkably, the subsidy coverage required to induce permanent entry is quite large for most of these firms: 29% and 18% of all "induceable" firms need subsidies above 40% and 50% of their R&D expenditures respectively to engage in R&D.

[INSERT TABLE 12]

Table 13 provides a breakdown by industries. The percentage of R&D firms varies widely across industries ranging from 4% in printing products to 54% is office and data processing machinery (see column 1). Most firms in low-tech industries need both entry and continuation subsidies but the percentage decreases for medium-tech and high-tech industries (see column 3). The percentage of firms with positive entry thresholds and negative continuation thresholds is remarkable in all industries and particularly in the medium-tech and high-tech ones (see column 4). This implies that there is room for increasing the percentage of R&D firms in all industries. Column (2) shows the maximum percentage of R&D firms that can be attained in every single industry by adding up the numbers of columns (4) and (5).

[INSERT TABLE 13]

6.3. Evaluation of R&D inducing subsidy policies

The third question we set out to address concerns the effectiveness of a policy aimed at inducing all the firms with positive entry thresholds and negative continuation thresholds to undertake R&D. To carry out this evaluation we assume that subsidies are granted at t=0 and then set equal to zero from t=1 onwards. Thus, all we need to do is to contrast the inducement costs, namely the total amount of subsidies granted at t=0, with the stream of R&D investments that are subsequently manifested.

In order to infer these inducement costs, it is convenient to express the subsidy thresholds in absolute terms rather than as a proportion of a firm's R&D expenditure. This can be easily achieved by multiplying the subsidy coverage threshold by the R&D threshold: $subsidy^E = \tilde{\rho}^E \tilde{x}^E$ and $subsidy^C = \tilde{\rho}^C \tilde{x}^C$. Then, the cost of inducing all the firms that only need entry subsidies is obtained by adding up the entry subsidies $(subsidy^E)$ of all firms with $\tilde{\rho}^E > 0 \; and \; \tilde{\rho}^C \leq 0$ that are not performing R&D yet[28]. We find that inducing all these firms (about 3,000 firms) would cost around €110 million. To obtain an idea as to whether these numbers make sense it might be helpful to know that in 2009 the CDTI (Spain's national agency of technology) spent €584 million on direct subsidies to finance 944 projects. In comparison with this benchmark €110 million seems a very low number.

There are several potential explanations for why we get such a small number. First, we are focusing on manufacturing firms while most subsidies from CDTI go to service sectors. Second, the average subsidy granted by CDTI in 2009 was notably larger than the average subsidy in our sample (€620,000 vs. €135,000). Third, firms identified as "induceable" have lower $\hat{x}^*$ (€138,000 vs. €349,000) and $\hat{\tilde{x}}^E$ (€165,000 vs. €256,000) than subsidized firms. This implies that subsidies aimed at inducing permanent R&D performance have to subsidize lower quantities than subsidies awarded to active R&D firms. In any case, we must admit that our estimate seems a lower bound of the true inducement costs.

The yearly R&D investments that would be triggered by this inducing policy from t=1 onward are estimated at €453 million. This implies that, considering an optimistic

---

[28] The observations are weighted following the steps mentioned above in order to obtain representative results for the whole population of Spanish manufacturing firms.

scenario, in which no induced firm would abandon R&D activities after t=1, the R&D stock generated would scale up to €2,500 million in 15 years and reach a steady level of almost €3,000 million in 35 years. Under a more pessimistic scenario, in which half of the induced firms would abandon R&D after t=1, the R&D stock generated would reach a steady level of €1,500 million in 20 years. This implies that the inducing policy would still be effective even if the inducement costs were tenfold the estimated ones and half of the induced firms failed to persist into R&D.

To sum up, "extensive" subsidies can be used to expand the share of R&D firms in the Spanish manufacture from 20% to 30% which would in turn lead to an increase of R&D intensity (understood as total R&D expenditures over total sales) from 0.64% to 0.74%. So going back to Figure 1, "extensive" subsidies have the potential for placing Spain somewhere between Italy and Ireland, but not further. [29] For Spain to reach the group of countries formed by France, the Netherlands, Austria and Belgium, more structural policies consisting in lowering the entry and continuation subsidy thresholds should be implemented. Recall from expression (8) that the thresholds can be lowered by improving demand conditions, lowering the marginal and the fixed costs and shortening the lag between R&D and profits.

7. Conclusion

Researchers and policymakers alike have paid little attention to subsidies as a tool for expanding the base of R&D performers. But it is likely that there are sunk entry costs associated to R&D that can be smoothed out by subsidies and hence justify the existence of "extensive" subsides. In this paper we have sought to contribute fresh evidence regarding the feasibility and efficiency of such subsidies.

We have framed our analysis around a dynamic model with sunk entry costs in which firms decide whether to start, continue or stop performing R&D on the grounds of the subsidy coverage they expect to receive. The main appeal of this framework is that a firm's optimal participation strategy can be characterized in terms of two subsidy

[29] Notice that this only a rough comparison as our analysis takes into account both intramural and extramural R&D while Figure 1 considers only intramural R&D.

thresholds characterizing R&D entry and continuation. The existence of two thresholds proves crucial as it allows us to detect which firms need subsidies to start but not to continue performing R&D. Recognizing the risky nature of R&D (which we have not done in this paper) would only reinforce the argument of the existence of entry-deterring sunk costs.

We are able to compute the subsidy thresholds from the estimates of a dynamic panel data type-2 tobit model with an R&D investment equation and an R&D participation equation. By including dynamics in the selection equation, we are able to estimate true state dependence, which is ultimately used to measure the distance between the two thresholds. The model is estimated for an unbalanced panel of about 2,000 Spanish manufacturing firms observed over a 12-year period.

We find that expected subsidies significantly affect both R&D expenditure and the decision to perform R&D. In addition, we conclude that R&D performance is true state dependent which leads to the existence of two subsidy thresholds, one that determines entry into R&D and one that assures continuation of R&D.

Using the estimated expected subsidy coverage thresholds we find that 25% of Spanish manufacturing firms need subsidies to start but not to continue doing R&D. Slightly more than half of the firms belonging to this group are already R&D performers, which means that the other half (10% of Spanish manufacturing firms) is still to be induced. Should they be induced, the proportion of R&D firms in the Spanish manufacture would increase by about one half (from 20% to 30%). We estimate that inducing this group of firms into R&D would cost €110 million, while the stream of R&D investments that this would give rise to would generate an R&D stock of €2,500 million over a 15-year period. This result emphasizes the importance of dynamic additionality, generally disregarded in analyses of subsidy effectiveness.

The findings offered by this paper call for a revision of the classical subsidy granting schemes. Subsidies have traditionally been awarded to consolidated R&D performers. However, agencies have shown a certain reluctance to award subsidies to reduce the entry costs for R&D beginners. This is mainly because they do not really know whether there is scope for using subsidies to induce entry into R&D, and they are unaware of the costs involved. This paper has confirmed that subsidies can be used to defray the sunk costs and encourage entry into R&D. Besides, the costs of inducing all these firms have been found to be relatively moderate compared with the R&D stock that would be generated.

Of course, subsidies aimed at inducing entry may well generate moral hazard problems when actually implemented. For instance, firms might be tempted to perform R&D during a period simply to receive the subsidy and, once the subsidy received, they cease their R&D activities. Similarly, they might over-invest in R&D so as to obtain larger subsidies. A solution might be to tie the provision of such funds to a commitment from the firms to invest similar amounts in R&D during the subsequent years. Only firms that intend to continue their R&D activities are likely to accept such a contract. Unquestionably, the optimal design of subsidies aimed at inducing sustained R&D merits careful consideration and constitutes a topic for future research.

Appendix A. Calculation of the expected subsidy coverage

Columns (1) and (2) of Table 3A report estimates of the parameter vectors λ_1 and λ_2 using an augmented version of the specification proposed in González et al. (2005). The dependent variable of the probit regression is a dummy with value one for successful subsidy applicants and value zero for rejected applicants (successful applicant dummy). The dependent variable of the OLS regression is the natural logarithm of the subsidy coverage for successful applicants.

Regarding the explanatory variables, the lagged endogenous variables are included to capture persistence in the probability of getting subsidies and in the subsidy coverage. There are some cases in which firms that received a subsidy at time t did not receive a subsidy at t-1. For this reason we complement the log of the subsidy in the OLS equation with a dummy variable taking value of one if the firm did not receive a subsidy at t-1.

We include as explanatory variables those considered in González et al. (2005) that public agencies may consider as critical when making their decisions: firm size, age, degree of technological sophistication, a dummy indicating whether the firm is a domestic exporter, a dummy denoting whether the firm has foreign capital and a dummy indicating whether the firm has market power. Finally, time, region and industry dummies are also included. We also include some explanatory variables not considered in González et al. (2005): a lagged R&D dummy to reflect the fact that regular R&D firms are more likely to get subsidies while R&D entrants are likely to be awarded larger subsidy shares, and the initial value of the lagged dependent variables, the R&D dummy and the "no subsidy dummy" to capture firms' unobservable heterogeneity. Some of these explanatory variables are considered as predetermined and are thus included with a lag, while others are assumed to be strictly exogenous.

The definitions and descriptive statistics of the variables used in the regressions are reported in tables 1A and 2A respectively. The results are reported in Table 3A. The probability of receiving a subsidy (column 1) is higher for applicants who received subsidies in the past, have experience in R&D, and are large and technologically sophisticated. Subsidy coverage (column 2) depends on the past coverage, and agencies appear to be more inclined to award large subsidies to R&D entrants, to small firms and to firms with market power. All the parameters of the initial conditions are significant.

Using the estimated $\hat{\lambda}_1$ and $\hat{\lambda}_2$ we calculate the expected subsidy coverage $\hat{\rho}_{it}^e$ as follows:

$$\hat{\rho}_{it}^e = \begin{cases} \Phi(z_{it}^{\rho}\hat{\lambda}_1)\exp(z_{it}^{\rho}\hat{\lambda}_2 + (1/2)\hat{\sigma}^2) & if\ ap_{it} = 1 \\ 0 & if\ ap_{it} = 0 \end{cases} \tag{1A}$$

The estimated expected subsidies have reasonable values. The average probability of receiving a subsidy among applicants is 68%, the average expected subsidy coverage conditional on its being granted is 31%, and the average expected subsidy is 2%. Only a small proportion of the expected conditional subsidy coverages take values higher than 100%. There are five observations for which the predicted unconditional expected subsidy coverage takes a value higher than 100%. For these five observations we replaced the predicted value by 99% to calculate $-\ln(1-\hat{\rho}_{it}^e)$.

In columns (3) and (5) we estimate parameter vectors λ_3 and λ_5 omitting the lagged R&D dummy and the initial conditions (that is, using the specification proposed in González et al., 2005). We calculate $\hat{\rho}_{it}^e\_1$ as follows:

$$\hat{\rho}_{it}^e\_1 = \begin{cases} \Phi(z_{it}^{\rho}\hat{\lambda}_3)\exp(z_{it}^{\rho}\hat{\lambda}_5 + (1/2)\hat{\sigma}^2) & if\ ap_{it} = 1 \\ 0 & if\ ap_{it} = 0 \end{cases} \tag{2A}$$

In column (4) we obtain the parameter vector λ_4 by estimating the probit for the whole sample (not just for subsidy applicants) and again using the same specification proposed in González et al. (2005). We calculate $\hat{\rho}_{it}^e\_2$ as in González et al. (2005):

$$\hat{\rho}_{it}^e\_2 = \Phi(z_{it}^{\rho}\hat{\lambda}_4)\exp(z_{it}^{\rho}\hat{\lambda}_5 + (1/2)\hat{\sigma}^2) \tag{3A}$$

Descriptive statistics of the different expected subsidy coverage measures are provided in Table 4A.

[INSERT TABLE 1A]
[INSERT TABLE 2A]
[INSERT TABLE 3A]
[INSERT TABLE 4A]

References

Aschhoff, B. (2010). “Who gets the money? The Dynamics of R&D project subsidies in Germany”, *Jahrbücher für Nationalökonomie und Statistik*, 230(5), 522-546.

Baldwin, R. (1989), “Sunk Costs Hysteresis”, NBER Working Paper No. 2911.

Blanes J.V. and Busom I. (2004). “Who participates in R&D subsidy programs? The case of Spanish manufacturing firms”, *Research Policy*, 33 (10), 1459-1476.

Clerides S.K., Lach, S. and Tybout, J.R. (1998), “Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico and Morocco”, *The Quarterly Journal of Economics*, 113, 903-947.

David, P., Hall, B.H., Toole, A.A., (2000), “Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence”, *Research Policy*, 29, 497-529.

Dixit, A. and Stiglitz, J. (1977), “Monopolistic Competition and Optimum Product Diversity”, *American Economic Review*, 67, 297-308.

García-Quevedo J. (2004). “Do public subsidies complement business R&D? A Meta-Analysis of the Econometric Evidence”, *Kyklos*, 57, 87-102.

Gónzalez, X. and Jaumandreu, J. (1998), “Threshold Effects in Product R&D Decisions: Theoretical Framework and Empirical Analysis”, FEDEA working papers, 45.

González, X., Jaumandreu, J. and Pazó, C. (2005), “Barriers to Innovation and Subsidy Effectiveness”, *Rand Journal of Economics,* 36 (4), 930-949.

González, X., and Pazó, C. (2008), “Do Public Subsidies Stimulate Private R&D Spending?”, *Research Policy,* 37, 371-389.

Heckman, JJ. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. In *Structural Analysis of Discrete Data with Econometric Applications*, Manski CF, McFadden D (eds). MIT Press: Cambridge, MA.

Huergo, E. and Trenado, M. (2010). "The application for and the awarding of low-interest credits to finance R&D projects", *Review of Industrial Organization*, 37; 237-259.

Klette, T.J., Moen, J., and Griliches, Z. (2000), "Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies", *Research Policy*, 29, 471–495.

Levin, R. and Reiss, P. (1988), "Cost-Reducing and Demand-Creating R&D with Spillovers", *Rand Journal of Economics*, 19 (4), 538-556.

Malerba, F. and Orsenigo, L. (1999), "Technological Entry, Exit and Survival: An Empirical Analysis of Patent Data", *Research Policy*, 28, 643-660.

Mansfield, E., Rapoport, J., Schnee, J., Wagner, S., Hamburger, M., 1971. Research and development in the modern corporation. W.W. Norton, New York.

Mañez, J., Rochina, M. E., Sanchis, A. and Sanchis, J.A. (2009), "The Role of Sunk Costs in the Decision to Invest in R&D", *The Journal of Industrial Economics*, 57 (4), 712-735.

Melitz, M. (2000), "Estimating Firm-Level Productivity in Differentiated Product Industries"; Harvard University.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data", *Econometrica*, 46, 69-85.

Nelson, F. (1977), "Censored Regression Models with Unobserved Stochastic Censored Thresholds." *Journal of Econometrics*, 6, 309–327.

Neyman, J. and Scott, E. (1948), "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, 1-32.

Organization for Economic Co-operation and Development (2002), *The Measurement of Scientific Technical Activities. Frascati Manual 2002. Proposed Standard Practice for Surveys of Research and Experimental Development*, Paris.

Peters, B. (2009), "Persistence of Innovation: Stylised Facts and Panel Data Evidence", *Journal of Technology Transfer*, 34(2), 226-243.

Ravenscraft, D. and Scherer, F.M. (1982). The lag structure of returns to research and development, *Applied Economics*, 14, 603-620.

Raymond, W., Mohnen, P., Palm, F. and Schim, S. (2010), "Persistence of Innovation in Dutch Manufacturing: Is it Spurious?", *Review of Economics and Statistics*, 92(3), 495–504.

Stewart, M. (2007), "The Interrelated Dynamics of Unemployment and Low-Wage Employment", *Journal of Applied Econometrics*, 22, 511-531.

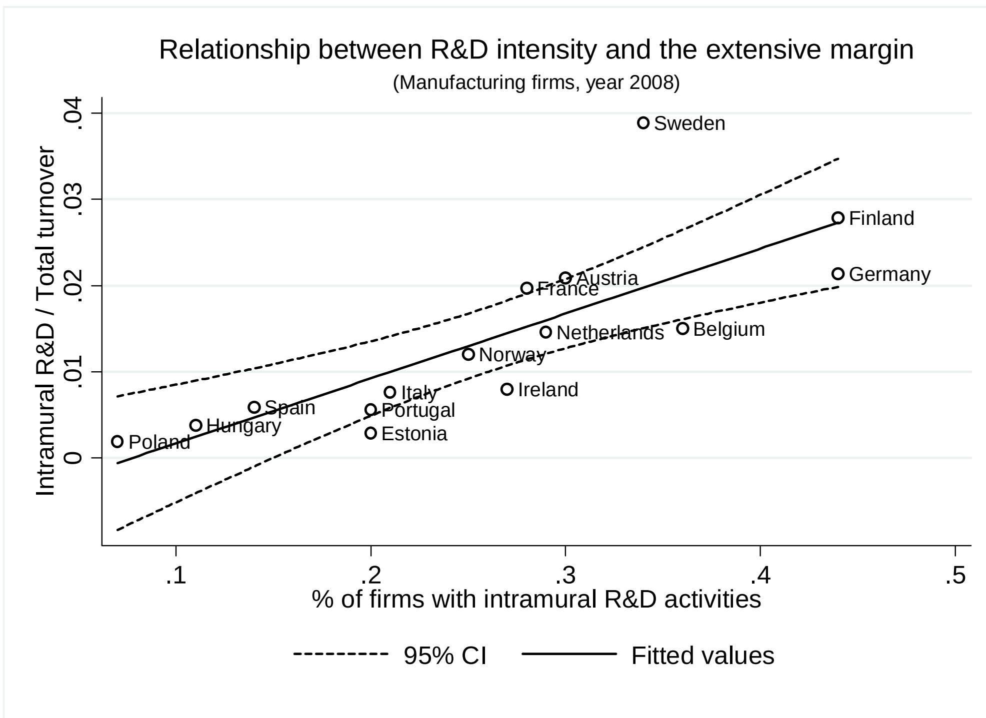
Sutton, J. (1991), *Sunk Costs and Market Structure*. MIT Press.

Takalo, T., Tanayama, T. and Toivanen, O. (2011), "Estimating the Benefits of Targeted Subsidies ", *Review of Economics and Statistics,* forthcoming.

Wooldridge, J. (2005), "Simple Solutions to the Initial Conditions Problem in Dynamic Nonlinear Panel Data Models with Unobserved Heterogeneity", *Journal of Applied Econometrics*, 20(1), 39-54.

FIGURES

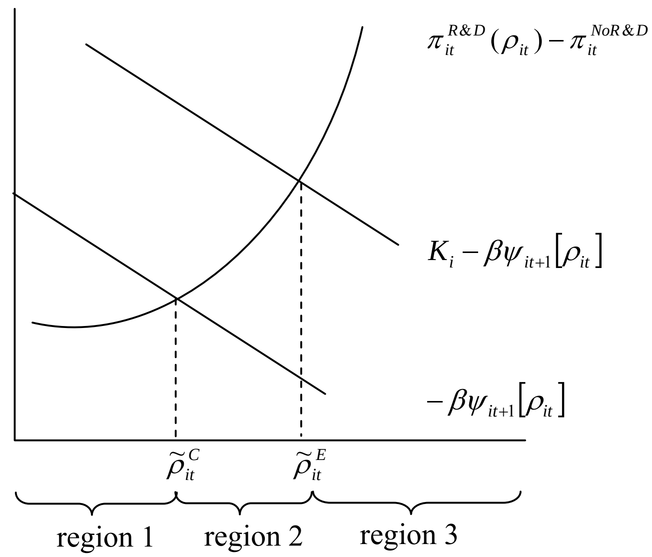
Figure 1



Source: Community Innovation Survey, 2008. Eurostat.

Figure 2

Representation of the indifference condition that defines the thresholds

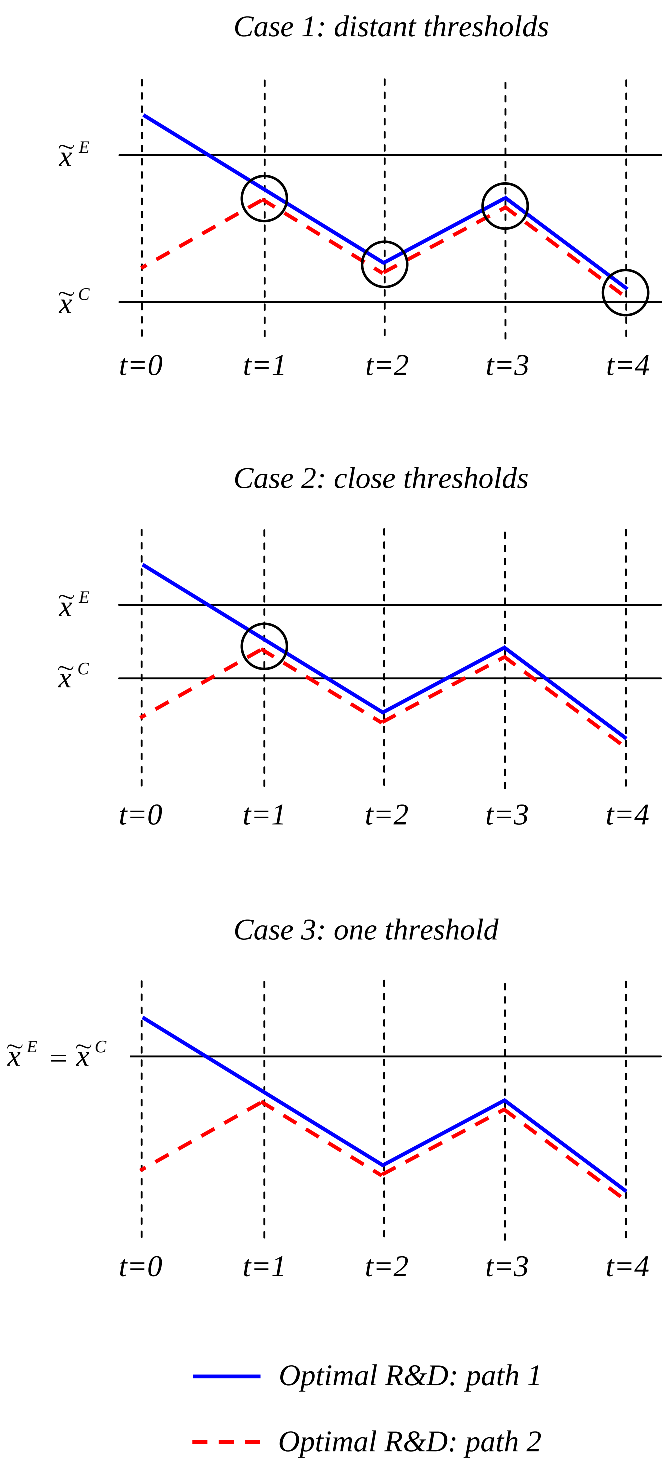


Notes: provided that the following condition is satisfied

$$\partial\left(\pi_{it}^{R\&D}(\tilde{\rho}_{it}) - \pi_{it}^{NoR\&D} + \beta\psi_{it+1}[\tilde{\rho}_{it}]\right)/\partial\rho_{it} > 0$$

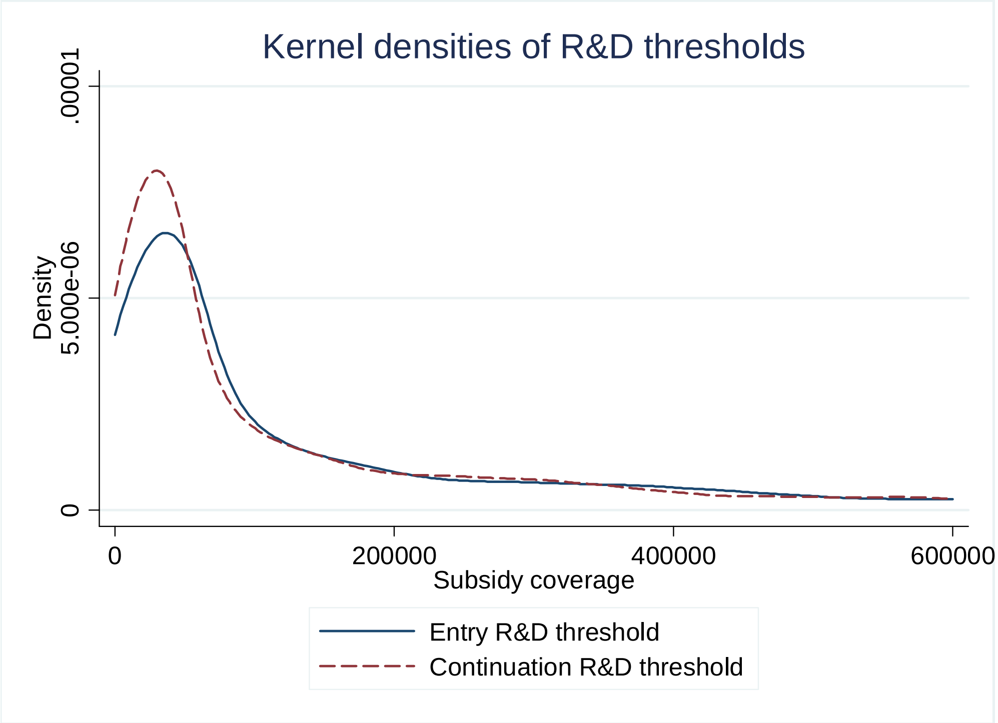
period t optimal entry-exit strategy can be depicted as in Figure 2. This figure shows that the thresholds define three regions into which subsidies can be differentiated. Region 1 contains all those subsidies for which a firm will not perform R&D regardless of its history. At the opposite end of the spectrum, region 3 contains all those subsidies for which the firm will perform R&D regardless of its history. Finally, region 2 identifies all those values of a subsidy for which a firm's previous status does matter. More specifically, a firm expecting to receive a subsidy that falls between the boundaries defined by region 2 will only perform R&D if it was already performing R&D in the previous period.

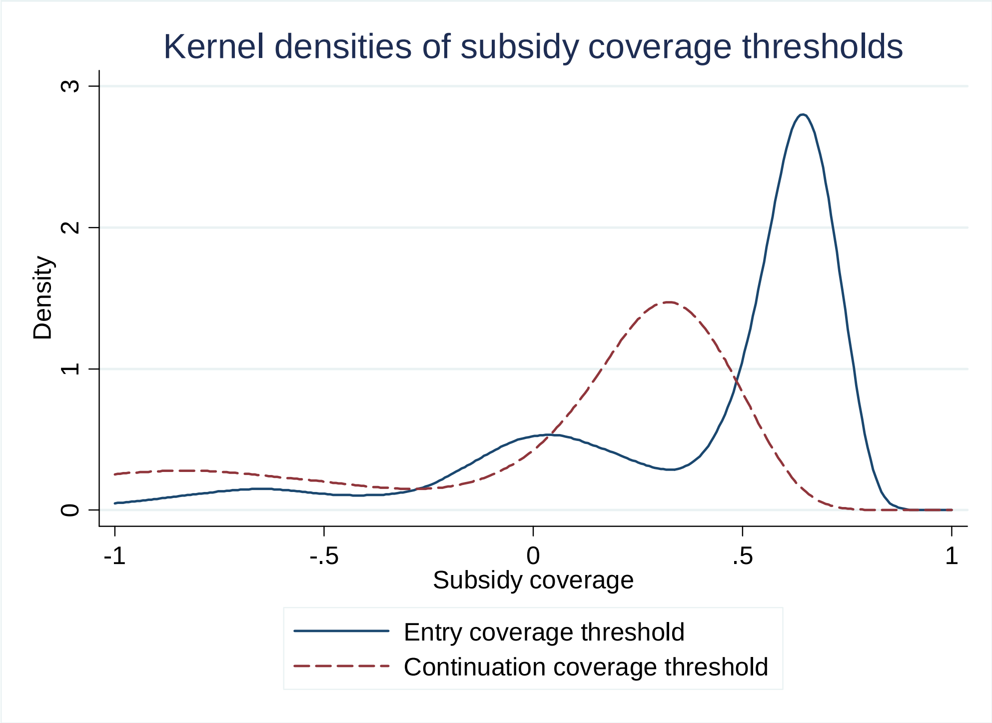
Figure 3. Relationship between the thresholds and true state dependence



Note: the circles ◯ identify the periods in which, under path 1, firms perform R&D because they were already performing R&D in the previous period.

Figure 4. Kernel densities of R&D and subsidy coverage thresholds





TABLES

Table 1. Composition of the panel

t	number of firms
2	366
3	444
4	430
5	135
6	195
7	80
8	104
9	168
10	138
11	80
12	142
13	339
Number of firms	2,621

Notes: this table shows the number of firms that are observed for each spell length.

Table 2. Successful applicants, rejected applicants and non applicants

	Number of firms	%
Successful applicants		
Continue in R&D	1,013	94
Enter into R&D	69	6
Fail to enter into R&D	0	0
Exit from R&D	0	0
Total	1,082	100
Rejected applicants		
Continue in R&D	319	63
Enter into R&D	44	9
Fail to enter into R&D	119	24
Exit from R&D	21	4
Total	503	100
Non applicants		
Continue in R&D	2,708	21
Enter into R&D	371	3
Fail to/do not want to enter into R&D	9,103	72
Exit from R&D	516	4
Total	12,698	100

Table 3. Percentage of R&D and subsidized firms by firm size

	Fewer than 200 workers			More than 200 workers		
	Firms with R&D (%)	Subsidized R&D firms (%)	Average subsidy share (%)	Firms with R&D (%)	Subsidized R&D firms (%)	Average subsidy share (%)
1998	19	16	26	73	25	19
1999	21	14	31	77	27	19
2000	21	18	34	74	29	25
2001	18	11	29	71	29	27
2002	19	14	34	72	27	23
2003	18	10	34	69	27	23
2004	18	13	37	70	27	24
2005	19	16	33	70	31	22
2006	19	16	33	70	33	27
2007	19	20	35	65	33	28
2008	19	23	37	65	34	35
2009	20	25	34	65	39	29

Notes: this table reports the percentage of R&D performers, the percentage of subsidized firms among R&D performers and average subsidies for subsidized firms in each year of the sample in a breakdown by size.

Table 4. R&D and subsidies by firm size and frequency of R&D performance

	Among all firms, % of		Firms granted subsidies at least one year, in %, among all	
	Stable R&D performers	Occasional R&D performers	Stable R&D performers	Occasional R&D performers
<20 workers	3	11	31	21
21-50	9	18	32	19
51-100	24	26	40	23
101-200	34	23	39	31
201-500	51	27	53	26
>500	66	19	56	41

Table 5: Variable definitions

Dependent variables

R&D expenditures: cost of intramural R&D activities and R&D contracted with external laboratories.

R&D dummy: dummy that takes the value one if R&D expenditure is positive.

Explanatory variables of interest

Expected subsidy coverage ($\hat{\rho}_{it}^{e}$): computed by equation (26). Product of the predicted probability of receiving a subsidy (estimated from subsidy applicants) and the expected value of the subsidy (estimated from successful subsidy applicants) for subsidy applicants, zero for non applicants. Two alternative measures $\hat{\rho}_{it\_}^{e}1$ and $\hat{\rho}_{it\_}^{e}2$ are experimented with (see appendix A for their construction).

Controls

Advertising/sales ratio: advertising expenditure over sales.

Average industry patents: yearly average number of patents registered by the firms in the same industry (excluding the patents registered by the firm), for a breakdown of manufacturing in 20 industries.

Average variable costs: total variable costs divided by nominal output (sales) so they really measure costs per unit revenue. Total variable costs are constructed as the sum of labour costs, intermediate input costs and subcontracted production costs.

Concentrated market: dummy variable that takes the value one if the firm reports that its main market consists of fewer than 10 competitors.

Expansive market: dummy variable that takes the value one if the firm reports that its main market is in expansion.

Foreign capital dummy: dummy that takes the value one if the firm has foreign capital.

Industry dummies: set of 20 industry dummies (NACE-09 classification). The first dummy (industry group 1: "meat industry") is used as the base category and is therefore excluded from the regressions.

Initial condition: dummy that takes value one if the firm performs R&D in the first year of the sample used for conducting the estimates and zero otherwise.

Market share: market share reported by the firms in its main market.

Quality controls: dummy variable that takes the value one if the firm carries out quality controls on a regular basis.

Recessive market: dummy variable that takes the value one if the firm reports that its main market is in recession.

Sales: total sales made by the firm.

Skilled labour: dummy that takes the value one if the firm possesses highly qualified workers (engineers and graduates).

Size dummies: set of six dummy variables, measuring size in terms of number of employees.

Subsidy applicant dummy (ap_{it}): dummy that takes the value one if the firm has applied for a subsidy (i.e., has received a subsidy or claims that has searched external funding without success).

Time dummies: set of 12 yearly dummy variables.

Table 6. Descriptive statistics

		standard deviation				
	mean	overall	between	within	min	max
Dependent variables						
Ln(R&D expenditures)	12.18	1.85	1.88	0.74	4.04	15.89
R&D dummy t	0.32	0.47	0.43	0.21	0	1
Explanatory variables						
R&D dummy t-1	0.32	0.47	0.43	0.22	0	1
R&D dummy 0	0.33	0.47	0.47	0	0	1
$-\ln(1-\hat{\rho}^{e}_{it})$	0.03	0.16	0.20	0.10	0	4.61
$-\ln(1-\hat{\rho}^{e}_{it-1})$	0.03	0.15	0.14	0.10	0	6.21
$-\ln(1-\hat{\rho}^{e}_{it-2})$	0.03	0.09	0.08	0.06	0	1.93
$m(-\ln(1-\hat{\rho}^{e}_{it}))$	0.03	0.13	0.20	0.00	0	4.61
$m(-\ln(1-\hat{\rho}^{e}_{it-1}))$	0.03	0.11	0.14	0.00	0	3.21
$m(-\ln(1-\hat{\rho}^{e}_{it-2}))$	0.03	0.07	0.08	0.00	0	1.10
Subsidy applicant dummy	0.11	0.31	0.28	0.19	0	1
Market share t-1	0.10	0.17	0.16	0.08	0	1
Concentrated market dummy t-1	0.53	0.50	0.43	0.29	0	1
Advertising sales ratio t-1	0.01	0.03	0.03	0.01	0	0.72
Average industry patents	0.24	0.66	0.32	0.57	0	12.56
Ln(Average variable costs t-1)	-0.09	0.14	0.11	0.10	-0.96	0.94
Recessive market t-1	0.19	0.40	0.30	0.31	0	1
Expansive market t-1	0.27	0.44	0.33	0.34	0	1
Foreign capital dummy	0.18	0.38	0.37	0.12	0	1
Quality controls	0.42	0.49	0.44	0.28	0	1
Skilled labour	0.64	0.48	0.45	0.19	0	1
21-50 workers	0.27	0.44	0.39	0.21	0	1
51-100 workers	0.10	0.31	0.26	0.17	0	1
101-200 workers	0.10	0.30	0.28	0.16	0	1
201-500 workers	0.17	0.37	0.35	0.15	0	1
>500 workers	0.08	0.26	0.26	0.09	0	1
Ln(Sales)	15.64	1.92	1.94	0.28	9.55	22.36

Notes: m( ) denotes the Mundlak mean of the variable in parentheses.

Table 7. Results of the dynamic panel data type-2 tobit

	(ln of) Optimal R&D expenditure		R&D decision		(ln of) R&D threshold	
	b_1 [a]	s.d. [b]	b_2 [a]	s.d. [b]	b_0 [a]	s.d. [b]
R&D dummy t-1			1.60	(0.05)***	(η) -0.20	(0.09)**
Dummy R&D 0	0.49	(0.07)***	1.71	(0.11)***	0.27	(0.11)**
$-\ln(1-\hat{\rho}_{it}^{e})$	(γ) 0.31	(0.10)***	(δ_0) 2.48	(0.70)***		
$m(-\ln(1-\hat{\rho}_{it}^{e}))$	0.26	(0.15)*	0.04	(0.33)	0.25	(0.16)
Applicant dummy			1.20	(0.14)***	-0.15	(0.08)**
Market share t-1	0.15	(0.11)	-0.04	(0.15)	0.16	(0.11)
Concentrated market dummy t-1	-0.03	(0.04)	0.05	(0.05)	-0.04	(0.04)
Advertising sales ratio t-1	2.37	(0.72)***	-0.23	(0.77)	2.40	(0.73)***
Average industry patents	-0.01	(0.03)	0.06	(0.04)	-0.02	(0.03)
ln(Average variable costs t-1)	-0.11	(0.14)	-0.51	(0.17)***	-0.05	(0.15)
Recessive market t-1	0.01	(0.05)	0.02	(0.06)	0.01	(0.05)
Extensive market t-1	-0.02	(0.04)	0.04	(0.05)	-0.03	(0.04)
Dummy foreign capital			0.03	(0.06)	0.00	(0.01)
Quality controls			0.10	(0.05)**	-0.01	(0.01)
Skilled labour			0.23	(0.07)***	-0.03	(0.02)*
21-50 workers	0.26	(0.09)***	0.14	(0.08)	-0.02	(0.01)
51-100 workers	0.58	(0.12)***	0.09	(0.11)	-0.01	(0.02)
101-200 workers	0.81	(0.12)***	0.00	(0.13)	0.00	(0.02)
201-500 workers	0.97	(0.14)***	0.21	(0.15)	-0.03	(0.02)
>500 workers	1.19	(0.17)***	-0.08	(0.20)	0.01	(0.03)
ln(Sales)	0.43	(0.03)***	0.21	(0.04)***	0.40	(0.03)***
Constant	2.91	(0.52)***	-6.20	(0.54)***	3.69	(0.61)***
Industry and time dummies	Yes		Yes		Yes	
ρ_{a1a2}	0.19	(0.04)***				
$\rho_{\varepsilon1\varepsilon2}$	-0.30	(0.04)***				
σ_{a1}	0.72	(0.05)***				
σ_{a2}			0.11	(0.05)**		
σ_{a0}					0.77	(0.11)***
$\sigma_{\varepsilon1}$	1.00	(0.01)***				
$\sigma_{\varepsilon2}=\gamma/\delta_0$			0.13	(0.05)**		
$\sigma_{\varepsilon0}$					0.97	(0.01)***
AME R&D dummy t-1			0.37			
Log likelihood			-9,482.94			
Number of observations	4,524		14,283		14,283	
Number of firms	1,104		2,621		2,621	

a) b1, b2 and b0 refer to the parameters of equations (13), (15) and (14) respectively. The coefficients of the threshold equation have been calculated as $b_0 = b_1 - \sigma_{\varepsilon2} * b_2$

b) Standard errors are shown in parentheses. The standard errors of the threshold equation have been calculated according to the delta method. ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Gaps indicate exclusion restrictions.

c) The AME of the R&D dummy at t-1 measures the average probability of doing R&D at t when y_{it-1} is fixed at 1 minus the average probability of doing R&D at t when y_{it-1} is fixed at 0, and it is evaluated at the average values of the covariates (see Stewart 2007).

Table 8. Robustness check 1: results of the dynamic panel data type-2 tobit using observations with $\rho_{it} < 0.60$ and $\rho_{it} < 0.75$

	All firms (1)	$\rho_{it} < 0.60$ (2)	$\rho_{it} < 0.75$ (3)
		Optimal R&D equation	
Dummy R&D 0	0.49 (0.07)***	0.39 (0.06)***	0.40 (0.06)***
$-\ln(1-\hat{\rho}_{it}^{e})$	0.31 (0.10)***	0.64 (0.13)***	0.58 (0.12)***
$m(-\ln(1-\hat{\rho}_{it}^{e}))$	0.26 (0.15)*	0.12 (0.17)	0.18 (0.16)
Constant	2.91 (0.52)***	2.19 (0.48)***	2.21 (0.49)***
Other variables	Yes	Yes	Yes
# of observations	4,524	4,379	4,446
# of firms	1,104	1,092	1,098
		Selection equation	
Dummy R&D t-1	1.60 (0.05)***	1.61 (0.05)***	1.61 (0.05)***
Dummy R&D 0	1.71 (0.11)***	1.69 (0.11)***	1.67 (0.11)***
$-\ln(1-\hat{\rho}_{it}^{e})$	2.48 (0.70)***	2.26 (0.72)**	2.38 (0.71)***
$m(-\ln(1-\hat{\rho}_{it}^{e}))$	0.04 (0.33)	0.05 (0.34)	0.03 (0.34)
Constant	-6.20 (0.54)***	-6.17 (0.54)***	-6.19 (0.54)***
Other variables	Yes	Yes	Yes
# of observations	14,283	14,138	14,205
# of firms	2,621	2,614	2,618
Log likelihood	-9,482.94	-9,266.18	-9,369.21

Notes: ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. The dependent variables are the natural logarithm of R&D expenditures and a dummy with value one if the firm performs R&D. Besides the shown coefficients the regressions also include all the controls included in Table 7. Column (1) reproduces the results shown in Table 7.

Table 9. Robustness check 2: results of the dynamic panel data type-2 tobit using $\hat{\rho}^{e}_{it\_}1$ and $\hat{\rho}^{e}_{it\_}2$

	$rho = \hat{\rho}^{e}_{it\_}1$	$rho = \hat{\rho}^{e}_{it\_}2$		
	(1)	(2)	(3)	(4)
		Optimal R&D equation		
Dummy R&D 0	0.38 (0.06) ***		0.58 (0.07) ***	0.46 (0.07) ***
$-\ln(1-rho)$	0.30 (0.09) ***	0.86 (0.14) ***	0.82 (0.13) ***	0.17 (0.16)
$m(-\ln(1-rho))$	0.67 (0.13) ***			2.05 (0.26) ***
Constant	2.15 (0.46) ***	2.61 (0.48) ***	2.59 (0.47) ***	1.87 (0.48) ***
Other variables	Yes	Yes	Yes	Yes
# of observations	4,524	4,524	4,524	4,524
# of firms	1,104	1,104	1,104	1,104
		Selection equation		
Dummy R&D t-1	1.59 (0.05) ***		1.60 (0.05) ***	1.60 (0.05) ***
Dummy R&D 0	1.66 (0.11) ***		1.87 (0.12) ***	1.77 (0.12) ***
$-\ln(1-rho)$	1.99 (0.80) **	3.61 (0.31) ***	0.95 (0.27) ***	-0.11 (0.30)
$m(-\ln(1-rho))$	0.59 (0.43)			4.33 (0.55) ***
Constant	-6.17 (0.54) ***	-8.05 (0.40) ***	-6.04 (0.52) ***	-5.74 (0.52) ***
Other variables	Yes	Yes	Yes	Yes
# of observations	14,283	14,283	14,283	14,283
# of firms	2,621	2,621	2,621	2,621
Log likelihood	-9,473.38	-11,179.48	-9,720.53	-9,666.40

Notes: ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. The dependent variables are the natural logarithm of R&D expenditures and a dummy with value one if the firm performs R&D. Besides the shown coefficients the regressions also include all the controls included in Table 7.

Table 10. Distribution of R&D and subsidy coverage thresholds

R&D thresholds					Subsidy thresholds				
	$\widetilde{x}^E$		$\widetilde{x}^C$			$\widetilde{\rho}^E$		$\widetilde{\rho}^C$	
	%	Cum.	%	Cum.		%	Cum.	%	Cum.
< 50,000	42	42	47	47	< 0	18	18	38	38
50,000-100,000	15	57	14	61	0 - 0.2	10	28	13	51
100,000-150,000	7	64	7	68	0.2 - 0.4	6	34	33	84
150,000-200,000	6	70	5	73	0.4 - 0.6	22	56	15	100
200,000-250,000	4	73	4	77	0.6 - 0.8	44	100	0	100
>200,000	27	100	23	100	0.8 - 1	0	100	0	100

Table 11. Classification of firms according to their dependence on subsidies

Groups of firms according to their dependence on subsidies		(1) Observations in the sample		(2) Firms in the sample		(3) Firms in the population	
		A	B	A	B	A	B
1	$\widetilde{\rho}^E > 0 \,\&\, \widetilde{\rho}^C > 0$	62	5	58	3	70	4
2	$\widetilde{\rho}^E \leq 0 \,\&\, \widetilde{\rho}^C \leq 0$	18	93	13	95	5	98
3	$\widetilde{\rho}^E > 0 \,\&\, \widetilde{\rho}^C \leq 0$	20	60	10	51	9	56
4	$\widetilde{\rho}^E > 0 \,\&\, \widetilde{\rho}^C \leq 0$, $\widetilde{\rho}^E \leq 0 \,\&\, \widetilde{\rho}^C \leq 0$			11	73	9	72
5	$\widetilde{\rho}^E > 0 \,\&\, \widetilde{\rho}^C \leq 0$, $\widetilde{\rho}^E > 0 \,\&\, \widetilde{\rho}^C > 0$			8	41	7	40

A) Proportion of observations or firms that fall into each group out of the total.
B) Proportion of firms in each group that perform R&D (the proportion is calculated with respect to the total number of observations in each group, not with respect to the total number of firms in the sample)

Table 12. Percentage of firms that can be permanently induced with each range of subsidy coverage (out of the firms that can be permanently induced and are not performing R&D yet)

Entry subsidy coverage (in %)	% of firms
10	8
20	26
30	19
40	29
50	18

Note: these numbers are an extrapolation for the whole manufacture

Table 13. R&D and subsidies' permanent inducement effects by industries

	Current % of R&D firms	Maximum % of R&D firms (4) + (5)	% of firms with $\tilde{\rho}^E > 0$ & $\tilde{\rho}^C > 0$	$\tilde{\rho}^E > 0$ & $\tilde{\rho}^C < 0$	$\tilde{\rho}^E < 0$ & $\tilde{\rho}^C < 0$
	(1)	(2)	(3)	(4)	(5)
Low technological regime industries					
Meats, meat preparation	10	14	86	12	2
Beverages	29	49	51	37	12
Textiles and clothing	17	20	80	14	6
Leather, leather and skin goods	20	21	79	17	4
Timber, wooden products	8	10	90	8	2
Printing products	4	4	96	3	1
Paper	12	17	83	14	3
Non-metallic mineral products	13	18	82	15	3
Metal products	16	21	79	17	4
Furniture	16	20	80	17	3
Other manufacturing products	7	10	90	10	0
Medium technological regime industries					
Food products and tobacco	13	14	86	10	4
Rubber and plastic products	22	26	74	18	8
Ferrous and non-ferrous metals	33	52	48	35	17
Agricultural and industrial machinery	40	48	52	33	15
Motor vehicles	32	39	61	29	10
High technological regime industries					
Chemical products	52	59	41	19	40
Office and data processing machinery	54	65	35	38	27
Electrical goods	39	43	57	25	18
Other transport equipment	40	44	56	29	15

Notes: these numbers are an extrapolation for the whole manufacture.

TABLES APPENDIX

Table 1A: Variable definition

Dependent variables

Subsidy coverage: ratio of total public subsidies to total R&D expenditure. Total R&D expenditure of the firm includes the cost of intramural R&D activities and payments for outside R&D contracts (this definition of R&D is consistent with the definition given in the Frascati Manual).

Subsidy dummy: dummy that takes the value one for successful subsidy applicants and zero for the remaining firms (rejected applicants and non-applicants).

Successful applicant dummy: dummy that takes the value one for successful subsidy applicants and zero for rejected subsidy applicants.

Explanatory variables

Age: firms' age.

Domestic exporter dummy: dummy variable that takes the value one if the firm is domestic (less than 50% of foreign capital) and has exported during the year.

Foreign capital dummy: dummy that takes the value one if the firm has foreign capital.

Industry dummies: set of 20 industry dummies (NACE-09 classification). The first dummy (industry group 1: "meat industry") is used as the base category and is therefore excluded from the regressions.

Market power dummy: dummy variable that takes the value one if the firm claims to have market power.

No subsidy dummy: dummy variable that takes the value one if the firm does not receive subsidies.

R&D dummy: dummy that takes the value one if R&D expenditure is positive.

Region dummies: set of 17 autonomous community (region) dummies.

Size: number of employees in the firm.

Technological sophistication: dummy variable that takes the value one if the firm uses automatic machines, or robot or CAD/CAM, or some combination of these procedures, multiplied by the ratio of engineers and university graduates to total personnel.

Time dummies: set of 12 yearly dummy variables.

Table 2A. Descriptive statistics of variables included in the subsidy regressions

	mean	standard deviation overall	between	within	min	max
Dependent variables						
Subsidy dummy	0.08	0.26	0.24	0.16	0	1
Subsidy coverage	0.02	0.10	0.09	0.07	0	1
ln(Subsidy coverage)	-0.14	0.57	0.48	0.37	-5.79	0
Explanatory variables						
R&D dummy $_{t-1}$	0.32	0.47	0.43	0.22	0	1
R&D dummy $_{0}$	0.33	0.47	0.47	0	0	1
Size $_{t-1}$	162	315	352	66	2	6,648
Age	25	19	19	6	1	169
Technological sophistication	0.07	0.12	0.12	0.06	0	1
Domestic exporter dummy $_{t-1}$	0.47	0.50	0.47	0.21	0	1
Foreign capital dummy	0.18	0.38	0.37	0.12	0	1
Market power dummy $_{t-1}$	0.31	0.46	0.41	0.24	0	1

The variable ln(Subsidy coverage) is the natural logarithm of subsidy coverage for subsidy coverages greater than zero and zero for subsidy shares equal to zero. Notice that the variable size takes values lower than 10 for some observations despite the fact that the ESEE only incorporates firms with more than 9 workers. This is because once incorporated in the survey some firms decrease below 9 employees.

Table 3A. Subsidy regressions used to calculate expected subsidy coverage

Sample used in the regressions:	Subsidy applicants	Subsidized firms	Subsidy applicants	All firms	Subsidized firms
Dependent variable:	Successful applicant dummy	ln(Subsidy coverage)	Successful applicant dummy	Subsidy dummy	ln(Subsidy coverage)
Parameters estimated:	λ_1	λ_2	λ_3	λ_4	λ_5
	(1)	(2)	(3)	(4)	(5)
Subsidy dummy $_{t-1}$	1.08 (0.11)***		1.57 (0.11)***	2.00 (0.06)***	
Subsidy dummy $_0$	1.31 (0.16)***				
ln(subsidy coverage) $_{t-1}$		0.24 (0.04)***			0.45 (0.04)***
ln(subsidy coverage) $_0$		0.55 (0.04)***			
No subsidy dummy $_{t-1}$		-0.56 (0.10)***			-0.83 (0.09)***
No subsidy dummy $_0$		-0.85 (0.11)***			
R&D dummy $_{t-1}$	0.30 (0.15)**	-0.42 (0.13)***			
R&D dummy $_0$	0.06 (0.15)	-0.29 (0.12)**			
Size $_{t-1}$	0.00 (0.00)*	-0.00 (0.00)**	0.00 (0.00)**	0.00 (0.00)***	-0.00 (0.00)**
Age	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)*	0.00 (0.00)**	-0.01 (0.00)***
Tech. sophistication	0.87 (0.38)**	0.16 (0.16)	1.18 (0.39)***	1.10 (0.15)***	-0.10 (0.18)
Domestic exporter $_{t-1}$	0.22 (0.14)	-0.11 (0.09)	0.28 (0.13)**	0.42 (0.05)***	-0.10 (0.10)
Foreign capital dummy	0.10 (0.17)	-0.13 (0.09)	0.10 (0.16)	0.22 (0.07)***	-0.26 (0.10)**
Market power dummy $_{t-1}$	0.02 (0.11)	0.12 (0.06)**	0.04 (0.10)	0.17 (0.05)***	-0.03 (0.06)
Constant	-0.21 (0.45)	0.06 (0.24)	0.11 (0.45)	-2.93 (0.20)***	-0.93 (0.23)***
Industry, region and time dummies	yes	yes	yes	yes	yes
σ		0.92			1.01
Estimation method	Probit	OLS	Probit	Probit	OLS
# of observations	1,585	1,082	1,585	14,278	1,082
# of firms	588	411	588	2,619	411
R2		0.39			0.26
Pseudo R2	0.44		0.36	0.45	

Notes: ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Firm-clustered-robust standard errors are shown in parentheses. The dependent variable in columns (1) and (3) is a dummy variable with value one for successful applicants and value zero for rejected applicants. The dependent variable in column (4) is a dummy variable with value one for successful applicants and value zero for the remaining firms in the sample (rejected applicants and non-applicants). The dependent variable in columns (2) and (5) is the natural logarithm of the subsidy coverage. 19 industry dummies, 16 region dummies and 12 year dummies have been included.

Table 4A. Descriptive statistics of the actual subsidy coverage and the different measures of expected subsidy coverage

	mean	standard deviation overall	standard deviation between	standard deviation within	min	max
	All firms in the sample (N=14,283)					
ρ_{it}	0.02	0.10	0.09	0.07	0	1
$\hat{\rho}^{e}_{it}$	0.02	0.09	0.09	0.05	0	0.99
$\hat{\rho}^{e}_{it\_}1$	0.02	0.09	0.08	0.06	0	1
$\hat{\rho}^{e}_{it\_}2$	0.02	0.07	0.06	0.04	0	0.85
	All applicants (N=1,585)					
ρ_{it}	0.19	0.25	0.23	0.15	0	1
$\hat{\rho}^{e}_{it}$	0.21	0.17	0.18	0.06	0.01	0.99
$\hat{\rho}^{e}_{it\_}1$	0.22	0.16	0.14	0.09	0.03	1
$\hat{\rho}^{e}_{it\_}2$	0.11	0.14	0.11	0.08	0	0.85
	Successful applicants (N=1,082)					
ρ_{it}	0.28	0.25	0.24	0.16	0	1
$\hat{\rho}^{e}_{it}$	0.26	0.19	0.19	0.07	0.02	0.99
$\hat{\rho}^{e}_{it\_}1$	0.26	0.17	0.15	0.10	0.03	1
$\hat{\rho}^{e}_{it\_}2$	0.15	0.14	0.12	0.09	0	0.85

The UNU-MERIT WORKING Paper Series

2012-01 *Maastricht reflections on innovation* by Luc Soete
2012-02 *A methodological survey of dynamic microsimulation models* by Jinjing Li and Cathal O'Donoghue
2012-03 *Evaluating binary alignment methods in microsimulation models* by Jinjing Li and Cathal O'Donoghue
2012-04 *Estimates of the value of patent rights in China* by Can Huang
2012-05 *The impact of malnutrition and post traumatic stress disorder on the performance of working memory in children* by Elise de Neubourg and Chris de Neubourg
2012-06 *Cross-national trends in permanent earnings inequality and earnings instability in Europe 1994*-2001 by Denisa Maria Sologon and Cathal O'Donoghue
2012-07 *Foreign aid transaction costs* by Frieda Vandeninden
2012-08 *A simulation of social pensions in Europe* by Frieda Vandeninden
2012-09 *The informal ICT sector and innovation processes in Senegal* by Almamy Konté and Mariama Ndong
2012-10 *The monkey on your back?! Hierarchical positions and their influence on participants' behaviour within communities of learning* by Martin Rehm, Wim Gijselaers and Mien Segers
2012-11 *Do Ak models really lack transitional dynamics?* by Yoseph Yilma Getachew
2012-12 *The co-evolution of organizational performance and emotional contagion* by R. Cowan, N. Jonard, and R.Weehuizen
2012-13 *"Surfeiting, the appetite may sicken": Entrepreneurship and the happiness of nations* by Wim Naudé, José Ernesto Amorós and Oscar Cristi
2012-14 *Social interactions and complex networks* by Daniel C. Opolot
2012-15 *New firm creation and failure: A matching approach* by Thomas Gries, Stefan Jungblut and Wim Naudé
2012-16 *Gains from child-centred Early Childhood Education: Evidence from a Dutch pilot programme* by Robert Bauchmüller
2012-17 *Highly skilled temporary return, technological change and Innovation: The Case of the TRQN Project in Afghanistan* by Melissa Siegel and Katie Kuschminder
2012-18 *New Technologies in remittances sending: Opportunities for mobile remittances in Africa* Melissa Siegel and Sonja Fransen
2012-19 *Implementation of cross-country migration surveys in conflict-affected settings: Lessons from the IS Academy survey in Burundi and Ethiopia* by Sonja Fransen, Katie Kuschminder and Melissa Siegel
2012-20 *International entrepreneurship and technological capabilities in the Middle East and North Africa* by Juliane Brach and Wim Naudé
2012-21 *Entrepreneurship, stages of development, and industrialization* by Zoltan J. Ács and Wim Naudé
2012-22 *Innovation strategies and employment in Latin American firms* by Gustavo Crespi and Pluvia Zuniga
2012-23 *An exploration of agricultural grassroots innovation in South Africa and implications for innovation indicator development* by Brigid Letty, Zanele Shezi and Maxwell Mudhara
2012-24 *Employment effect of innovation: microdata evidence from Bangladesh and Pakistan* by Abdul Waheed

2012-25 *Open innovation, contracts, and intellectual property rights: an exploratory empirical study* by John Hagedoorn and Ann-Kristin Ridder
2012-26 *Remittances provide resilience against disasters in Africa* by Wim Naudé and Henri Bezuidenhout
2012-27 *Entrepreneurship and economic development: Theory, evidence and policy* by Wim Naudé
2012-28 *Whom to target - an obvious choice?* by Esther Schüring and Franziska Gassmann
2012-29 *Sunk costs, extensive R&D subsidies and permanent inducement effects* by Pere Arqué-Castells and Pierre Mohnen